Name:_____Haojuan He_____

ECON 140

ID:_____3033721461_____ *Fall 2020*

Homework 8

1. In this homework you will investigate the relationship between income per capita and democracy. Is higher *GDP* per capita associated with a greater degree of democracy? To explore this question, we'll use the data in file ***democracy.csv.*** This is a panel data set for the years 1960-2000 that includes the variables described in Table 1:

Table 1.

DATA DESCRIPTION, FILE: democracy.csv

Variable name	Variable Description
fhpolr	Freedom House Political Rights Index - measure of democracy. Normalized to vary between 0 and 1.
lrgdppc	Natural log of GDP per capita (chain weighted 1996 prices).
laborshare	Labor share of value added.
lpop	Natural log of total population
socialism	Dummy variable = 1 if political regime is socialism.
code_numeric	Country numerical code.
year	Year of measurement (1960 – 2000 in 5-year increments).

Note: Questions begin on next page.

(a) Table 2 reports the results of four regressions, one in each column, each with *fhpolr* as the dependent variable. Column (1) asks you to run a regression with lagged values of *fhpolr* and *lrgdppc* as the independent variables. Column (2) asks you to add year dummies. Column (3) asks you to add both year and country dummies. Finally, column (4) asks you to add further controls in addition to year and country dummies. Estimate the regressions and fill in the values (type the entries in) in the table below.

Table 2. Regression Results

Method	**(1) OLS**	**(2) FE**	**(3) FE**	**(4) FE**
Dependent variable:	$fhpolr_t$	$fhpolr_t$	$fhpolr_t$	$fhpolr_t$
$fhpolr_{t-1}$	0.5982 SE(0.039)	0.605 SE(0.040655)	0.147211 SE(0.062567)	0.142283 SE(0.063029)
$lrgdppc_{t-1}$	0.1210 SE(0.0140)	0.115386 SE(0.014436)	0.065786 SE (0.070633)	0.052051 SE(0.073183)
$laborshare_{t-1}$				0.1666617 SE(0.187661)
$lpop_{t-1}$				- 0.041594 SE(0.124298)
$socialism$				(None)
Constant	-0,76652 SE(0.103)			
Country dummies?	*No*	*No*	*Yes*	*Yes*
Year dummies?	*No*	*Yes*	*Yes*	*Yes*
R^2 *(Unadjusted)*	0.682	0.68026	0.023783	0.026955
R^2(adjusted)	0.680	0.67408	-0.31469	-0.32005
RSS	14.9035	14.06642	9.4164	9.3858
N	370	370	370	370
Standard errors		Clustered	Clustered	Clustered

(b) Provide an interpretation of the regression coefficients for both included covariates in regression (1).

For every 1 unit increase in fnpolr.lag in the last 5-year period, this should result in a 0.5982 unit increase in the current Freedom House Political Rights Index. And for every 1% increase in lrgdppc.lag in the previous 5-year period, this should results in a 0.1210 unit increase in the current Freedom House Political Rights Index.

(c) Why does the researcher need to exclude one of the year fixed effects in regression (2)?

The researcher need to exclude one of the year fixed effects in regression in order to avoid the Dummy variable trap. Because the regression did not exclude the intercept, not exclude one of the year fixed effects will results in perfect collinearity with dummy variables. This will cause our regression to fail as the model would not be able to separate the independent variable's effect on the dependent variable.

(d) Discuss some factors that might be accounted for by country and time fixed effects.

Country fixed effects: entity factors that affect the dependent variable (democracy index) over time

- the political systems of a country does not change easily over time but vary a lot among different countries

Time fixed effects: time factors that affect every entity the same but change over different time period

- Global socio-economic change such as the current pandemic and the 5G internet development that affect every countries the same way

(e) Comparing regressions (1) and (2), does the association between GDP and democracy change when you add year fixed effects? State and test this hypothesis.

(f) Comparing regression (2) and (3), does the association between GDP and democracy when you add country fixed effects in addition to year fixed effects? State and test this hypothesis.

(g) Interpret the coefficient on *Socialism* in regression (4).

From the data we can conclude that socialism does not change over time. As a result, in our plm regression, socialism is treated as an entity fixed effect. When plm looks for variability in this parameter, it treat this variable as a fixed effect and thus does not display it in the regression result.

(h) Comparing regressions (3) and (4), does adding controls change the results?

hw 8.

e.

⇒ let null hypothesis be association btw GDP & democracy does not change when add year fix effect

⇒ for regression

⇒ $fhpolr = \beta_0 + \beta_1 \cdot fhpolr\_lag + \beta_2 \cdot lrgdppc\_lag + \gamma_1 DY_1 + \gamma_2 DY_2 + \dots \gamma_n DY_n$

⇒ N = # unique year − 1

⇒ N = 5

⇒ H_0 is $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5 = 0$

⇒ alternative hypothesis = any $\gamma_i \neq 0$

⇒ assume H_0 is true.

$$\Rightarrow F = \frac{(RSS_{R.} - RSS_{u.R.})/\#q}{RSS_{u.R.}/(n-k-1)} = \frac{(R^2_{R.} - R^2_{u.R.})/q}{(1-R^2_{u.R.})/(n-k-1)}$$

$$= \frac{(14.90353 - 14.06640)/5}{14.06640/362}$$

$$= 4.30872$$

⇒ n = 370 & $q_i = 5$

$\Rightarrow n-k-1 = 370 - (9-1) = 362$

$\Rightarrow$ critical $F_{5\%(6,362)} = 2.24$

$\Rightarrow 4.30872 > 2.12 \quad (F > F_\alpha)$

$\Rightarrow$ we reject the null hypothesis that year does not have an impact on GDP at all

$\Rightarrow$ thus, the year does have an impact on GDP so the association btw GDP & democracy change when add year fix effect.

h.

$\Rightarrow$ let our null hypothesis be that adding additional control does not change the result

$\Rightarrow$ denote k_1 is the coefficient for $lpop_{t-1}$

$\Rightarrow$ denote k_2 is the coefficient for socialism

$\Rightarrow$ denote k_3 is the coefficient for $labshare_t$

$\Rightarrow H_0$ is $k_1 = k_2 = k_3 = 0$

$\Rightarrow$ alternative hypothesis H_1 is one of the k_1, k_2, k_3

β not zero

⇒ assume H_0 is true

⇒ $F = \dfrac{(RSS_{R.} - RSS_{u.R.})/\#q}{RSS_{u.R.}/(n-k-1)} = \dfrac{(R^2_{R.} - R^2_{u.R.})/q}{(1-R^2_{u.R.})/(n-k-1)}$

⇒ here we use chisquare for panel data

⇒ chisquare = 0.8865667

⇒ $\mathbb{P}(\text{chisquare}_a > \text{chisquare}') \simeq 0.6419$

⇒ 0.6419 > 5%

⇒ thus, we fail to reject our null hypothesis that adding control does not change the result & the coefficient for those controls variable are jointly insignificant

⇒ thus, comparing regression (3) & (4), adding controls does not change the results.

f.

⇒ let null hypothesis be association btw GDP & democracy does not change when add additional country fixed effect

⇒ f-

⇒ for regression

⇒ $fhpolr = \beta_0 + \beta_1 \cdot fhpolr\_lag + \beta_2 \cdot lrgdppc\_lag$

$+ \gamma_1 DY_1 + \gamma_2 DY_2 + \ldots\ldots \gamma_n DY_n$

$+ \alpha_1 DC_1 + \ldots\ldots + \alpha_{88} DC_{88}$

⇒ #89 different unique countries

⇒ we have #88 dummy variable

⇒ H_0 is $\alpha_1 = \alpha_2 = \alpha_3 \ldots\ldots = \alpha_{88} = 0$

⇒ alternative hypothesis = any $\alpha_i \neq 0$

⇒ assume H_0 is true.

⇒ $F = \dfrac{(RSS_{R.} - RSS_{u.R.})/\#q}{RSS_{u.R.}/(n-k-1)} = \dfrac{(R^2_{R.} - R^2_{u.R.})/q}{(1-R^2_{u.R.})/(n-k-1)}$

$= 1.7168$

⇒ $n = 370$ & $q = 88$

⇒ $n-k-1 = 370 - (89-1) = 282$

⇒ critical $F_{5\%\,(88,\,282)} = 1.61$

⇒ $1.7168 > 1.61$ $(F > F_\alpha)$

⇒ we fail to reject the null hypothesis that country does not have an impact on GDP at all and the coefficient for those controls variable are

jointly insignificant

⇒ thus, comparing regression (2) & (3), adding country as fix effect. does not change the results.

In [27]:
```
install.packages("plm")
install.packages("lmtest")
install.packages("xtable")
install.packages("plyr")
install.packages("car")
library(xtable)
library(broom)
library(knitr)
library(plm)
library(lmtest)
library(tidyverse)
library(plyr)
library(car)

library(lmtest)

getwd()
data <- read.csv("democracy.csv", header = TRUE, sep = ",")

# lag calculation
data$fhpolr.lag <- ifelse(data$year == 1960, NA, lag(data$fhpolr))
data$lrgdppc.lag <- ifelse(data$year == 1960, NA, lag(data$lrgdppc))
data$laborshare.lag <- ifelse(data$year == 1960, NA, lag(data$laborshare))
data$lpop.lag <- ifelse(data$year == 1960, NA, lag(data$lpop))
clean_data <- drop_na(data)
```

```
Updating HTML index of packages in '.Library'
Making 'packages.html' ... done
Updating HTML index of packages in '.Library'
Making 'packages.html' ... done
Updating HTML index of packages in '.Library'
Making 'packages.html' ... done
Updating HTML index of packages in '.Library'
Making 'packages.html' ... done
Updating HTML index of packages in '.Library'
Making 'packages.html' ... done
Loading required package: carData

Attaching package: ‘car’

The following object is masked from ‘package:dplyr’:

    recode

The following object is masked from ‘package:purrr’:

    some


'/Users/haojuanhe/ECON140'
```

In [94]:
```
#clean_data
```

In [29]:
```
getwd()
data <- read.csv("democracy.csv", header = TRUE, sep = ",")

# lag calculation
data$fhpolr.lag <- ifelse(data$year == 1960, NA, lag(data$fhpolr))
data$lrgdppc.lag <- ifelse(data$year == 1960, NA, lag(data$lrgdppc))
data$laborshare.lag <- ifelse(data$year == 1960, NA, lag(data$laborshare))
data$lpop.lag <- ifelse(data$year == 1960, NA, lag(data$lpop))
clean_data <- drop_na(data)
```

'/Users/haojuanhe/ECON140'

In [96]:
```
head(clean_data)
```

	fhpolr	lrgdppc	laborshare	lpop	socialism	code_numeric	year	fhpolr.lag	lrgdppc.
49	0.8333333	9.202840	0.1989480	10.167740	0	6	1975	0.1666667	9.133
50	0.1666667	9.271142	0.1957906	10.243310	0	6	1980	0.8333333	9.202
51	0.8333333	9.065316	0.1852767	10.319070	0	6	1985	0.1666667	9.271
74	1.0000000	9.420744	0.5134989	9.340316	0	9	1965	1.0000000	9.277
75	1.0000000	9.603759	0.5442176	9.434044	0	9	1970	1.0000000	9.420
76	1.0000000	9.657485	0.5442806	9.539141	0	9	1975	1.0000000	9.603

In [31]:
```
#data$fhpolr.lag
```

In [32]:
```
fit_regular <- lm(fhpolr ~ fhpolr.lag + lrgdppc.lag, data = clean_data)
```

In [86]:
```
summary(fit_regular)
deviance(fit_regular)
```

```
Call:
lm(formula = fhpolr ~ fhpolr.lag + lrgdppc.lag, data = clean_data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.85719 -0.09009  0.00965  0.09350  0.74330

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.76652    0.10275  -7.460 6.32e-13 ***
fhpolr.lag   0.59819    0.03862  15.490  < 2e-16 ***
lrgdppc.lag  0.12100    0.01380   8.768  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2015 on 367 degrees of freedom
Multiple R-squared:  0.6822,    Adjusted R-squared:  0.6805
F-statistic: 393.9 on 2 and 367 DF,  p-value: < 2.2e-16
```

14.9035332340887

In [58]:
```
democracy.panel.year <- pdata.frame(clean_data, index=c("year"), stringsAsF
fit_year <- plm(fhpolr ~ fhpolr.lag + lrgdppc.lag + factor(year), data = de
```

In [87]:
```
summary(fit_year)
deviance(fit_year)
```

```
Oneway (individual) effect Within Model

Call:
plm(formula = fhpolr ~ fhpolr.lag + lrgdppc.lag + factor(year),
    data = democracy.panel.year, model = "within")

Unbalanced Panel: n = 6, T = 41-76, N = 370

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-0.8209093 -0.0958779 -0.0070831  0.0918557  0.7294974

Coefficients:
            Estimate Std. Error t-value  Pr(>|t|)
fhpolr.lag  0.605475   0.040655 14.8931 < 2.2e-16 ***
lrgdppc.lag 0.115386   0.014436  7.9929 1.785e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    43.994
Residual Sum of Squares: 14.066
R-Squared:      0.68026
Adj. R-Squared: 0.67408
F-statistic: 385.088 on 2 and 362 DF, p-value: < 2.22e-16
```

14.0664028193483

In [90]:
```
ocracy.panel.both <- pdata.frame(clean_data, index=c("year","code_numeric"),
_yearcountry <- plm(fhpolr ~ fhpolr.lag + lrgdppc.lag, data = democracy.pane
mary(fit_yearcountry)
iance(fit_yearcountry)
```

```
Twoways effects Within Model

Call:
plm(formula = fhpolr ~ fhpolr.lag + lrgdppc.lag, data = democracy.panel.b
oth,
    effect = "twoways", model = "within")

Unbalanced Panel: n = 6, T = 41-76, N = 370

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-0.5342452 -0.0536701 -0.0041979  0.0644677  0.5372875

Coefficients:
            Estimate Std. Error t-value Pr(>|t|)
fhpolr.lag  0.147211   0.062567  2.3528  0.01934 *
lrgdppc.lag 0.065786   0.070633  0.9314  0.35248
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    9.6458
Residual Sum of Squares: 9.4164
R-Squared:      0.023783
Adj. R-Squared: -0.31469
F-statistic: 3.33764 on 2 and 274 DF, p-value: 0.036971


9.41639584499784
```

In [88]:
```
fit_all <- plm(fhpolr ~ fhpolr.lag + lrgdppc.lag + labor share.lag +lpop.lag
summary(fit_all)
deviance(fit_all)
```

```
Twoways effects Within Model

Call:
plm(formula = fhpolr ~ fhpolr.lag + lrgdppc.lag + labor share.lag +
    lpop.lag + socialism, data = democracy.panel.both, effect = "twoway
s",
    model = "within")

Unbalanced Panel: n = 6, T = 41-76, N = 370

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-0.5375663 -0.0597981 -0.0052803  0.0630133  0.5423724

Coefficients:
                Estimate Std. Error t-value Pr(>|t|)
fhpolr.lag      0.142283   0.063029  2.2574  0.02477 *
lrgdppc.lag     0.052051   0.073183  0.7112  0.47754
labor share.lag  0.166617   0.187661  0.8879  0.37540
lpop.lag       -0.041594   0.124298 -0.3346  0.73816
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    9.6458
Residual Sum of Squares: 9.3858
R-Squared:      0.026955
Adj. R-Squared: -0.32005
F-statistic: 1.88368 on 4 and 272 DF, p-value: 0.11351

9.3858034141479
```

In [89]:
```
fit_regular_y <- lm(fhpolr ~ fhpolr.lag + lrgdppc.lag + factor(year), data
summary(fit_regular_y)
deviance(fit_regular_y)
```

```
Call:
lm(formula = fhpolr ~ fhpolr.lag + lrgdppc.lag + factor(year),
    data = clean_data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.82091 -0.09588 -0.00708  0.09186  0.72950

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      -0.65783    0.10430  -6.307 8.28e-10 ***
fhpolr.lag        0.60547    0.04065  14.893  < 2e-16 ***
lrgdppc.lag       0.11539    0.01444   7.993 1.79e-14 ***
factor(year)1970 -0.14657    0.03861  -3.796 0.000172 ***
factor(year)1975 -0.10464    0.03833  -2.730 0.006636 **
factor(year)1980 -0.03527    0.03872  -0.911 0.362924
factor(year)1985 -0.03594    0.03885  -0.925 0.355549
factor(year)1990 -0.06253    0.04380  -1.428 0.154258
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1971 on 362 degrees of freedom
Multiple R-squared:    0.7,     Adjusted R-squared:  0.6942
F-statistic: 120.7 on 7 and 362 DF,  p-value: < 2.2e-16
```

14.0664028193483

In [74]:
```
linearHypothesis(fit_regular_y, c("factor(year)1970=0", "factor(year)1975=0
```

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
367	14.90353	NA	NA	NA	NA
362	14.06640	5	0.8371304	4.308724	0.0008054684

In [76]:
```
f_stats1 <- ((14.90353-14.06640)/5)/(14.06640/362)
f_stats1
```

4.30872234544731

In [83]:
```
linearHypothesis(fit_all,c('laborshare.lag=0','lpop.lag =0'))
```

Res.Df	Df	Chisq	Pr(>Chisq)
274	NA	NA	NA
272	2	0.8865667	0.6419253

In [92]:
```
arcountry_p <- plm(fhpolr ~ fhpolr.lag + lrgdppc.lag, data = democracy.panel
(fit_yearcountry, fit_yearcountry_p)
```

```
        F test for twoways effects

data:  fhpolr ~ fhpolr.lag + lrgdppc.lag
F = 1.7168, df1 = 93, df2 = 274, p-value = 0.0004217
alternative hypothesis: significant effects
```

In [ ]:
```

```