# Problemset-2

Ester Agasha

2023-10-04

```
#load and run the required libraries
library(haven)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggplot2)
library(plm)
```

```
##
## Attaching package: 'plm'
##
## The following objects are masked from 'package:dplyr':
##
##     between, lag, lead
```

```
#Question 1
#uplooad the WDI_FDI dataset
wdi_fdi_data <-read_dta("WDI_FDI_data.dta")
head(wdi_fdi_data)
```

```
## # A tibble: 6 x 16
##   countryname countrycode imfcode region regioncode  year   pop gdp_gr gdp_pc_gr
##   <chr>       <chr>         <dbl> <chr>  <chr>      <dbl> <dbl>  <dbl>     <dbl>
## 1 United Sta~ USA             111 North~ NAC         1950    NA     NA        NA
## 2 United Sta~ USA             111 North~ NAC         1951    NA     NA        NA
## 3 United Sta~ USA             111 North~ NAC         1952    NA     NA        NA
## 4 United Sta~ USA             111 North~ NAC         1953    NA     NA        NA
## 5 United Sta~ USA             111 North~ NAC         1954    NA     NA        NA
## 6 United Sta~ USA             111 North~ NAC         1955    NA     NA        NA
## # i 7 more variables: fdi_net_usdol <dbl>, fdi_pcgdp <dbl>,
## #   fdi_netin_usdol <dbl>, fdi_netout_pcgdp <dbl>, fdi_netout_usdol <dbl>,
## #   inflation_cpi <dbl>, inflation_gdpdefl <dbl>
```

```r
tail(wdi_fdi_data)
```

```
## # A tibble: 6 x 16
##   countryname countrycode imfcode region          regioncode  year    pop gdp_gr
##   <chr>       <chr>         <dbl> <chr>           <chr>      <dbl>  <dbl>  <dbl>
## 1 Romania     ROU             968 Europe & Centr~ ECS         2010 2.02e7  -2.81
## 2 Romania     ROU             968 Europe & Centr~ ECS         2011 2.01e7   2.03
## 3 Romania     ROU             968 Europe & Centr~ ECS         2012 2.01e7   1.24
## 4 Romania     ROU             968 Europe & Centr~ ECS         2013 2.00e7   3.53
## 5 Romania     ROU             968 Europe & Centr~ ECS         2014 1.99e7   3.08
## 6 Romania     ROU             968 Europe & Centr~ ECS         2015 1.98e7   3.97
## # i 8 more variables: gdp_pc_gr <dbl>, fdi_net_usdol <dbl>, fdi_pcgdp <dbl>,
## #   fdi_netin_usdol <dbl>, fdi_netout_pcgdp <dbl>, fdi_netout_usdol <dbl>,
## #   inflation_cpi <dbl>, inflation_gdpdefl <dbl>
```

```r
#setting up the dataset in a panel structure format
panel_data <- pdata.frame(wdi_fdi_data, index = c("imfcode", "year"))
```

```r
#Question 1.a)
# Pooled OLS regression with lagged FDI variables
pooled_ols <- plm(gdp_pc_gr ~ fdi_pcgdp + lag(fdi_pcgdp) + lag(fdi_pcgdp, 2) + year, data = panel_data,
# Summary of regression results
summary(pooled_ols)
```

```
## Pooling Model
##
## Call:
## plm(formula = gdp_pc_gr ~ fdi_pcgdp + lag(fdi_pcgdp) + lag(fdi_pcgdp,
##     2) + year, data = panel_data, model = "pooling")
##
## Unbalanced Panel: n = 189, T = 2-44, N = 6159
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -54.861671  -2.324176   0.059128   2.375742 131.892707
##
## Coefficients:
##                      Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)        3.5605052  0.7439963  4.7856 1.744e-06 ***
## fdi_pcgdp          0.0160366  0.0071462  2.2441 0.0248640 *
## lag(fdi_pcgdp)     0.0247921  0.0078839  3.1446 0.0016708 **
## lag(fdi_pcgdp, 2)  0.0085675  0.0072136  1.1877 0.2350041
## year1973          -1.3754715  1.0388351 -1.3241 0.1855353
## year1974          -0.6644175  1.0191890 -0.6519 0.5144850
## year1975          -3.1920584  1.0228779 -3.1207 0.0018128 **
## year1976           0.3896638  0.9988684  0.3901 0.6964723
## year1977          -1.8507305  0.9710108 -1.9060 0.0566992 .
## year1978          -1.1711181  0.9570868 -1.2236 0.2211398
## year1979          -1.2389838  0.9338429 -1.3268 0.1846382
## year1980          -2.9589683  0.9256087 -3.1968 0.0013968 **
## year1981          -3.2625178  0.9209991 -3.5424 0.0003995 ***
## year1982          -4.5494196  0.9152333 -4.9708 6.850e-07 ***
```

```
## year1983          -4.5312872  0.9110982 -4.9734 6.757e-07 ***
## year1984          -2.4532938  0.9072077 -2.7042 0.0068653 **
## year1985          -2.3531829  0.9059489 -2.5975 0.0094136 **
## year1986          -2.3392173  0.9022986 -2.5925 0.0095506 **
## year1987          -2.3891134  0.8976548 -2.6615 0.0077996 **
## year1988          -1.0932806  0.8943842 -1.2224 0.2216097
## year1989          -2.1379116  0.8932992 -2.3933 0.0167287 *
## year1990          -2.1765585  0.8912242 -2.4422 0.0146257 *
## year1991          -2.6941043  0.8912115 -3.0230 0.0025135 **
## year1992          -2.6440105  0.8862168 -2.9835 0.0028612 **
## year1993          -3.1310927  0.8852523 -3.5369 0.0004078 ***
## year1994          -2.8543019  0.8772189 -3.2538 0.0011449 **
## year1995          -1.3884976  0.8724153 -1.5916 0.1115363
## year1996          -0.7764194  0.8693733 -0.8931 0.3718498
## year1997           0.0756204  0.8623478  0.0877 0.9301249
## year1998          -1.9870552  0.8617994 -2.3057 0.0211604 *
## year1999          -2.2434031  0.8605463 -2.6070 0.0091574 **
## year2000          -1.0025283  0.8592914 -1.1667 0.2433803
## year2001          -1.9610228  0.8587084 -2.2837 0.0224240 *
## year2002          -1.8516665  0.8568531 -2.1610 0.0307334 *
## year2003          -0.9359697  0.8545203 -1.0953 0.2734214
## year2004           0.7568807  0.8517866  0.8886 0.3742637
## year2005           0.0718021  0.8520192  0.0843 0.9328423
## year2006           0.6487935  0.8514755  0.7620 0.4461110
## year2007           0.3252249  0.8518818  0.3818 0.7026435
## year2008          -1.7555846  0.8523823 -2.0596 0.0394770 *
## year2009          -5.4475403  0.8515523 -6.3972 1.700e-10 ***
## year2010          -1.0308791  0.8515335 -1.2106 0.2260898
## year2011          -0.9885122  0.8519642 -1.1603 0.2459825
## year2012          -1.8609674  0.8528323 -2.1821 0.0291399 *
## year2013          -1.8819778  0.8528527 -2.2067 0.0273731 *
## year2014          -1.9582652  0.8520747 -2.2982 0.0215822 *
## year2015          -2.4454766  0.8522923 -2.8693 0.0041280 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    206430
## Residual Sum of Squares: 192790
## R-Squared:      0.06605
## Adj. R-Squared: 0.059021
## F-statistic: 9.39664 on 46 and 6112 DF, p-value: < 2.22e-16
```

```
# Dummy variable estimation with lagged FDI variables
dummy_estimation <- plm(gdp_pc_gr ~ fdi_pcgdp + lag(fdi_pcgdp) + lag(fdi_pcgdp, 2) + factor(imfcode) + y
# Summary of regression results
summary(dummy_estimation)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = gdp_pc_gr ~ fdi_pcgdp + lag(fdi_pcgdp) + lag(fdi_pcgdp,
##     2) + factor(imfcode) + year, data = panel_data, model = "within")
##
## Unbalanced Panel: n = 189, T = 2-44, N = 6159
```

```
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -49.96315  -2.03334   0.12867   2.17473 123.26212
##
## Coefficients:
##                     Estimate Std. Error t-value  Pr(>|t|)
## fdi_pcgdp          0.0099945  0.0069460  1.4389 0.1502353
## lag(fdi_pcgdp)     0.0204025  0.0075476  2.7032 0.0068878 **
## lag(fdi_pcgdp, 2)  0.0026271  0.0069975  0.3754 0.7073472
## year1973          -1.4235370  0.9867486 -1.4427 0.1491708
## year1974          -0.8055272  0.9686056 -0.8316 0.4056480
## year1975          -3.3121232  0.9719909 -3.4076 0.0006598 ***
## year1976           0.2459840  0.9497856  0.2590 0.7956528
## year1977          -2.0772123  0.9240905 -2.2478 0.0246227 *
## year1978          -1.4446243  0.9112303 -1.5854 0.1129389
## year1979          -1.5260524  0.8898192 -1.7150 0.0863949 .
## year1980          -3.2319119  0.8822557 -3.6632 0.0002512 ***
## year1981          -3.5279442  0.8780138 -4.0181 5.940e-05 ***
## year1982          -4.8407447  0.8727336 -5.5466 3.038e-08 ***
## year1983          -4.9378680  0.8689498 -5.6826 1.390e-08 ***
## year1984          -2.9347316  0.8654069 -3.3912 0.0007005 ***
## year1985          -2.8286451  0.8642581 -3.2729 0.0010705 **
## year1986          -2.8054399  0.8609299 -3.2586 0.0011259 **
## year1987          -2.8576086  0.8566857 -3.3357 0.0008562 ***
## year1988          -1.5753543  0.8537018 -1.8453 0.0650406 .
## year1989          -2.5948883  0.8526935 -3.0432 0.0023513 **
## year1990          -2.6997281  0.8509419 -3.1726 0.0015183 **
## year1991          -3.1946882  0.8512385 -3.7530 0.0001764 ***
## year1992          -3.1621740  0.8467384 -3.7345 0.0001898 ***
## year1993          -3.6670273  0.8460107 -4.3345 1.485e-05 ***
## year1994          -3.4947089  0.8387416 -4.1666 3.136e-05 ***
## year1995          -2.0629264  0.8342611 -2.4728 0.0134353 *
## year1996          -1.3998494  0.8313993 -1.6837 0.0922871 .
## year1997          -0.6263210  0.8250939 -0.7591 0.4478286
## year1998          -2.6539567  0.8246201 -3.2184 0.0012960 **
## year1999          -2.9163831  0.8235112 -3.5414 0.0004011 ***
## year2000          -1.6759811  0.8223882 -2.0379 0.0415999 *
## year2001          -2.6200270  0.8217830 -3.1882 0.0014390 **
## year2002          -2.5430716  0.8201453 -3.1008 0.0019393 **
## year2003          -1.6061506  0.8181104 -1.9632 0.0496645 *
## year2004           0.0498073  0.8156905  0.0611 0.9513123
## year2005          -0.6139164  0.8159003 -0.7524 0.4518161
## year2006          -0.0205215  0.8152972 -0.0252 0.9799198
## year2007          -0.3192769  0.8157416 -0.3914 0.6955196
## year2008          -2.3983355  0.8162217 -2.9383 0.0033125 **
## year2009          -6.1713895  0.8154797 -7.5678 4.376e-14 ***
## year2010          -1.7902198  0.8153933 -2.1955 0.0281643 *
## year2011          -1.7706206  0.8157445 -2.1706 0.0300042 *
## year2012          -2.6466173  0.8165392 -3.2413 0.0011966 **
## year2013          -2.6809465  0.8165504 -3.2833 0.0010321 **
## year2014          -2.7878186  0.8160799 -3.4161 0.0006395 ***
## year2015          -3.3008316  0.8163694 -4.0433 5.337e-05 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    180330
## Residual Sum of Squares: 168490
## R-Squared:      0.065651
## Adj. R-Squared: 0.028744
## F-statistic: 9.04874 on 46 and 5924 DF, p-value: < 2.22e-16
```

```
#Question 1.b)
# Run the fixed effects (fe) model
fe_model <- plm(gdp_pc_gr ~ fdi_pcgdp + lag(fdi_pcgdp) + lag(fdi_pcgdp, 2) + year, data = panel_data, m
summary(fe_model)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = gdp_pc_gr ~ fdi_pcgdp + lag(fdi_pcgdp) + lag(fdi_pcgdp,
##     2) + year, data = panel_data, model = "within")
##
## Unbalanced Panel: n = 189, T = 2-44, N = 6159
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -49.96315  -2.03334   0.12867   2.17473 123.26212
##
## Coefficients:
##                      Estimate Std. Error t-value  Pr(>|t|)
## fdi_pcgdp           0.0099945  0.0069460  1.4389 0.1502353
## lag(fdi_pcgdp)      0.0204025  0.0075476  2.7032 0.0068878 **
## lag(fdi_pcgdp, 2)   0.0026271  0.0069975  0.3754 0.7073472
## year1973           -1.4235370  0.9867486 -1.4427 0.1491708
## year1974           -0.8055272  0.9686056 -0.8316 0.4056480
## year1975           -3.3121232  0.9719909 -3.4076 0.0006598 ***
## year1976            0.2459840  0.9497856  0.2590 0.7956528
## year1977           -2.0772123  0.9240905 -2.2478 0.0246227 *
## year1978           -1.4446243  0.9112303 -1.5854 0.1129389
## year1979           -1.5260524  0.8898192 -1.7150 0.0863949 .
## year1980           -3.2319119  0.8822557 -3.6632 0.0002512 ***
## year1981           -3.5279442  0.8780138 -4.0181 5.940e-05 ***
## year1982           -4.8407447  0.8727336 -5.5466 3.038e-08 ***
## year1983           -4.9378680  0.8689498 -5.6826 1.390e-08 ***
## year1984           -2.9347316  0.8654069 -3.3912 0.0007005 ***
## year1985           -2.8286451  0.8642581 -3.2729 0.0010705 **
## year1986           -2.8054399  0.8609299 -3.2586 0.0011259 **
## year1987           -2.8576086  0.8566857 -3.3357 0.0008562 ***
## year1988           -1.5753543  0.8537018 -1.8453 0.0650406 .
## year1989           -2.5948883  0.8526935 -3.0432 0.0023513 **
## year1990           -2.6997281  0.8509419 -3.1726 0.0015183 **
## year1991           -3.1946882  0.8512385 -3.7530 0.0001764 ***
## year1992           -3.1621740  0.8467384 -3.7345 0.0001898 ***
## year1993           -3.6670273  0.8460107 -4.3345 1.485e-05 ***
## year1994           -3.4947089  0.8387416 -4.1666 3.136e-05 ***
## year1995           -2.0629264  0.8342611 -2.4728 0.0134353 *
## year1996           -1.3998494  0.8313993 -1.6837 0.0922871 .
```

```
## year1997          -0.6263210  0.8250939 -0.7591 0.4478286
## year1998          -2.6539567  0.8246201 -3.2184 0.0012960 **
## year1999          -2.9163831  0.8235112 -3.5414 0.0004011 ***
## year2000          -1.6759811  0.8223882 -2.0379 0.0415999 *
## year2001          -2.6200270  0.8217830 -3.1882 0.0014390 **
## year2002          -2.5430716  0.8201453 -3.1008 0.0019393 **
## year2003          -1.6061506  0.8181104 -1.9632 0.0496645 *
## year2004           0.0498073  0.8156905  0.0611 0.9513123
## year2005          -0.6139164  0.8159003 -0.7524 0.4518161
## year2006          -0.0205215  0.8152972 -0.0252 0.9799198
## year2007          -0.3192769  0.8157416 -0.3914 0.6955196
## year2008          -2.3983355  0.8162217 -2.9383 0.0033125 **
## year2009          -6.1713895  0.8154797 -7.5678 4.376e-14 ***
## year2010          -1.7902198  0.8153933 -2.1955 0.0281643 *
## year2011          -1.7706206  0.8157445 -2.1706 0.0300042 *
## year2012          -2.6466173  0.8165392 -3.2413 0.0011966 **
## year2013          -2.6809465  0.8165504 -3.2833 0.0010321 **
## year2014          -2.7878186  0.8160799 -3.4161 0.0006395 ***
## year2015          -3.3008316  0.8163694 -4.0433 5.337e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    180330
## Residual Sum of Squares: 168490
## R-Squared:      0.065651
## Adj. R-Squared: 0.028744
## F-statistic: 9.04874 on 46 and 5924 DF, p-value: < 2.22e-16
```

```
#Run the random effects (re) model
re_model <- plm(gdp_pc_gr ~ fdi_pcgdp + lag(fdi_pcgdp) + lag(fdi_pcgdp, 2), data = panel_data, model =
summary(re_model)
```

```
## Oneway (individual) effect Random Effect Model
##    (Swamy-Arora's transformation)
##
## Call:
## plm(formula = gdp_pc_gr ~ fdi_pcgdp + lag(fdi_pcgdp) + lag(fdi_pcgdp,
##     2), data = panel_data, model = "random")
##
## Unbalanced Panel: n = 189, T = 2-44, N = 6159
##
## Effects:
##                  var std.dev share
## idiosyncratic 30.064   5.483 0.909
## individual     3.013   1.736 0.091
## theta:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0873  0.5065  0.5486  0.5263  0.5701  0.5701
##
## Residuals:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -52.815  -2.068   0.173  -0.039   2.245 129.638
##
## Coefficients:
```

```
##                      Estimate Std. Error z-value  Pr(>|z|)
## (Intercept)         2.0058937  0.1498978 13.3817 < 2.2e-16 ***
## fdi_pcgdp           0.0208036  0.0070539  2.9492  0.003186 **
## lag(fdi_pcgdp)      0.0230242  0.0077187  2.9829  0.002855 **
## lag(fdi_pcgdp, 2) -0.0010732  0.0071025 -0.1511  0.879894
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    187950
## Residual Sum of Squares: 186110
## R-Squared:      0.010068
## Adj. R-Squared: 0.009586
## Chisq: 40.1755 on 3 DF, p-value: 9.7801e-09
```

```
#Run the hausman test to choose between fixed effects (fe) and random effects (re)
hausman_test <- phtest(fe_model, re_model)
# Print Hausman test results
print(hausman_test)
```

```
##
##  Hausman Test
##
## data:  gdp_pc_gr ~ fdi_pcgdp + lag(fdi_pcgdp) + lag(fdi_pcgdp, 2) +  ...
## chisq = 16.686, df = 3, p-value = 0.0008201
## alternative hypothesis: one model is inconsistent
```

```
#Question 1.c)
# Install the required packages
library(plm)
library(car)
```

```
## 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
```

```
# Run the fixed effects (fe) regression model with lagged variables
fe_model <- plm(gdp_pc_gr ~ lag(gdp_pc_gr) + lag(gdp_pc_gr, 2) + lag(fdi_pcgdp) + lag(fdi_pcgdp, 2) + y
# Run the f-test for joint significance of lagged FDI coefficients
ftest_fdi <- linearHypothesis(fe_model, c("lag(fdi_pcgdp) = 0", "lag(fdi_pcgdp, 2) = 0"))

# F-test for joint significance of lagged GDP coefficients
ftest_gdp <- linearHypothesis(fe_model, c("lag(gdp_pc_gr) = 0", "lag(gdp_pc_gr, 2) = 0"))
#view the fe results
summary(fe_model)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = gdp_pc_gr ~ lag(gdp_pc_gr) + lag(gdp_pc_gr, 2) +
##     lag(fdi_pcgdp) + lag(fdi_pcgdp, 2) + year, data = panel_data,
##     model = "within")
##
## Unbalanced Panel: n = 188, T = 6-44, N = 6122
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -78.56954  -1.83342   0.12469   2.00311 111.71615
##
## Coefficients:
##                      Estimate Std. Error t-value  Pr(>|t|)
## lag(gdp_pc_gr)     0.2485455  0.0133838 18.5706 < 2.2e-16 ***
## lag(gdp_pc_gr, 2)  0.0363408  0.0124827  2.9113 0.0036129 **
## lag(fdi_pcgdp)     0.0223539  0.0068652  3.2561 0.0011359 **
## lag(fdi_pcgdp, 2) -0.0024247  0.0068875 -0.3520 0.7248141
## year1973          -1.5661995  0.9778562 -1.6017 0.1092830
## year1974          -0.3466940  0.9564691 -0.3625 0.7170118
## year1975          -3.2558826  0.9602046 -3.3908 0.0007014 ***
## year1976           0.7588131  0.9504970  0.7983 0.4247095
## year1977          -2.5186561  0.9269899 -2.7170 0.0066064 **
## year1978          -1.0745868  0.9027992 -1.1903 0.2339831
## year1979          -1.3761006  0.8853033 -1.5544 0.1201468
## year1980          -2.8686704  0.8728807 -3.2864 0.0010206 **
## year1981          -2.8679622  0.8720591 -3.2887 0.0010123 **
## year1982          -3.8358377  0.8704164 -4.4069 1.067e-05 ***
## year1983          -3.7678909  0.8637184 -4.3624 1.308e-05 ***
## year1984          -1.7258786  0.8623818 -2.0013 0.0454067 *
## year1985          -2.1152751  0.8593995 -2.4613 0.0138704 *
## year1986          -2.2016006  0.8542309 -2.5773 0.0099820 **
## year1987          -2.2347767  0.8500494 -2.6290 0.0085860 **
## year1988          -0.9563823  0.8481032 -1.1277 0.2595044
## year1989          -2.2840499  0.8460061 -2.6998 0.0069579 **
## year1990          -2.1624637  0.8445647 -2.5604 0.0104785 *
## year1991          -3.0673022  0.8433710 -3.6370 0.0002782 ***
## year1992          -2.4532603  0.8432254 -2.9094 0.0036351 **
## year1993          -2.9571178  0.8399906 -3.5204 0.0004341 ***
## year1994          -2.5100719  0.8339062 -3.0100 0.0026234 **
## year1995          -1.1812694  0.8298684 -1.4234 0.1546611
## year1996          -0.9926524  0.8279985 -1.1989 0.2306315
## year1997          -0.4823576  0.8252079 -0.5845 0.5588871
## year1998          -2.6467515  0.8165596 -3.2413 0.0011963 **
## year1999          -2.4278289  0.8160441 -2.9751 0.0029406 **
## year2000          -1.0446516  0.8158954 -1.2804 0.2004639
## year2001          -2.2681656  0.8163771 -2.7783 0.0054812 **
## year2002          -2.0194195  0.8136561 -2.4819 0.0130958 *
## year2003          -1.0335800  0.8124920 -1.2721 0.2033839
## year2004           0.4567191  0.8096274  0.5641 0.5727006
## year2005          -0.7104101  0.8088423 -0.8783 0.3798141
## year2006           0.0058080  0.8070444  0.0072 0.9942582
## year2007          -0.4177376  0.8079378 -0.5170 0.6051464
```

```
## year2008          -2.4795157  0.8085208 -3.0667 0.0021740 **
## year2009          -5.7311682  0.8083938 -7.0896 1.503e-12 ***
## year2010          -0.3268873  0.8107962 -0.4032 0.6868391
## year2011          -1.5928847  0.8123982 -1.9607 0.0499589 *
## year2012          -2.2900477  0.8091718 -2.8301 0.0046689 **
## year2013          -2.1618744  0.8096734 -2.6701 0.0076047 **
## year2014          -2.1948035  0.8095619 -2.7111 0.0067255 **
## year2015          -2.6721050  0.8102036 -3.2981 0.0009793 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    187440
## Residual Sum of Squares: 164200
## R-Squared:      0.12399
## Adj. R-Squared: 0.089169
## F-statistic: 17.7284 on 47 and 5887 DF, p-value: < 2.22e-16
```

```
print(ftest_fdi)
```

```
## Linear hypothesis test
##
## Hypothesis:
## lag(fdi_pcgdp) = 0
## lag(fdi_pcgdp, 2) = 0
##
## Model 1: restricted model
## Model 2: gdp_pc_gr ~ lag(gdp_pc_gr) + lag(gdp_pc_gr, 2) + lag(fdi_pcgdp) +
##     lag(fdi_pcgdp, 2) + year
##
##   Res.Df Df Chisq Pr(>Chisq)
## 1   5889
## 2   5887  2 12.61   0.001827 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
print(ftest_gdp)
```

```
## Linear hypothesis test
##
## Hypothesis:
## lag(gdp_pc_gr) = 0
## lag(gdp_pc_gr, 2) = 0
##
## Model 1: restricted model
## Model 2: gdp_pc_gr ~ lag(gdp_pc_gr) + lag(gdp_pc_gr, 2) + lag(fdi_pcgdp) +
##     lag(fdi_pcgdp, 2) + year
##
##   Res.Df Df  Chisq Pr(>Chisq)
## 1   5889
## 2   5887  2 413.76  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#Joint significance test for the lagged variables
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```r
joint_test <- plmtest(fe_model, effect = "individual", type = "bp")
joint_test
```

```
##
##  Lagrange Multiplier Test - (Breusch-Pagan)
##
## data:  gdp_pc_gr ~ lag(gdp_pc_gr) + lag(gdp_pc_gr, 2) + lag(fdi_pcgdp) +  ...
## chisq = 23.55, df = 1, p-value = 1.217e-06
## alternative hypothesis: significant effects
```

```r
#Question 1. d)
#Load and run other neccessary libraries
library(plm)
#Run the regression with a two-way fixed effects using reghdfe
hdfe_reg <- plm(gdp_pc_gr ~ fdi_pcgdp + lag(fdi_pcgdp) + lag(fdi_pcgdp, 2), data = panel_data, within =
summary(hdfe_reg)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = gdp_pc_gr ~ fdi_pcgdp + lag(fdi_pcgdp) + lag(fdi_pcgdp,
##     2), data = panel_data, within = TRUE)
##
## Unbalanced Panel: n = 189, T = 2-44, N = 6159
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -50.73880  -1.96223   0.23637   2.25323 125.31860
##
## Coefficients:
##                      Estimate Std. Error t-value Pr(>|t|)
## fdi_pcgdp          0.0190539  0.0070816  2.6906 0.007152 **
## lag(fdi_pcgdp)     0.0216989  0.0077147  2.8127 0.004929 **
## lag(fdi_pcgdp, 2) -0.0029939  0.0071270 -0.4201 0.674446
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    180330
## Residual Sum of Squares: 179390
```

```
## R-Squared:      0.0051808
## Adj. R-Squared: -0.026663
## F-statistic: 10.3584 on 3 and 5967 DF, p-value: 8.5126e-07
```

```
#Question 2. a)
#Install the necessary packages
library(broom)
library(haven)
library(psych)
```

```
##
## Attaching package: 'psych'
```

```
## The following object is masked from 'package:car':
##
##     logit
```

```
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
library(irr)
```

```
## Loading required package: lpSolve
```

```
#load the woodstove data
file_path <- "C:/Users/agash/OneDrive/Desktop/Econometrics_Panel_Datasets/woodstove.dta"
woodstove <- read_dta(file_path)

#subset the data for the baseline year 2008
baseline_data <- subset(woodstove, year == 2008)

# List of dependent variables
dependent_variables <- c("educ", "weaknessdays", "diarrheadays", "coughdays",
                         "mucusdays", "redeyedays", "backpaindays", "faintdays", "feverdays")

#initialize vaiables to store ICC values and their corresponding var names
icc_values <- numeric ()
highest_icc_var <- ""
lowest_icc_var <- ""
highest_icc <- -Inf
lowest_icc <- Inf

#Looping through dep.vars and calculating ICC
for (var in dependent_variables) {
  #Calculate ICC using ANOVA
  icc_result <- summary(aov(get(var) ~ hh_id, data = baseline_data))

  #Extract ICC
  icc_value <- icc_result[[1]]$`Sum Sq`[1] / icc_result[[1]]$`Sum Sq`[2]
  #Uplating the higest and lowest ICC and their corresponding variable names
  if (icc_value > highest_icc) {
```

```
    highest_icc <- icc_value
    highest_icc_var <- var
  }
  if (icc_value < lowest_icc) {
    lowest_icc <- icc_value
    lowest_icc_var <- var
  }
  #Append ICC to the vector
  icc_values <- c(icc_values, icc_value)
}
#print results
```

```
cat("Highest ICC variable:", highest_icc_var, "\n")
```

```
## Highest ICC variable: backpaindays
```

```
cat("Highest ICC variable:", highest_icc, "\n")
```

```
## Highest ICC variable: 0.00702732
```

```
cat("Lowest ICC value:", lowest_icc, "\n")
```

```
## Lowest ICC value: 4.02575e-05
```

```
cat("Lowest ICC variable:", lowest_icc_var, "\n")
```

```
## Lowest ICC variable: diarrheadays
```

```
#Question 2. b)
# run Pooled OLS
pooled_ols <- lm(repcoughing ~ educ + female + age + age2, data = woodstove)
summary (pooled_ols)
```

```
##
## Call:
## lm(formula = repcoughing ~ educ + female + age + age2, data = woodstove)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7518 -0.2806 -0.2404  0.5509  0.8814
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.443e-01  2.132e-02  11.457  < 2e-16 ***
## educ        -1.054e-02  2.928e-03  -3.599 0.000325 ***
## female       3.021e-02  1.637e-02   1.846 0.065059 .
## age          2.579e-04  1.449e-03   0.178 0.858723
## age2         6.303e-05  2.102e-05   2.998 0.002737 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.4471 on 3040 degrees of freedom
##   (1140 observations deleted due to missingness)
## Multiple R-squared:  0.04638,    Adjusted R-squared:  0.04513
## F-statistic: 36.97 on 4 and 3040 DF,  p-value: < 2.2e-16
```

```
# run fixed effects (fe)
library(plm)
fixed_effects <- plm(repcoughing ~ educ + female + age + age2, data = woodstove, model = "within", inde
```

```
## Warning in pdata.frame(data, index): duplicate couples (id-time) in resulting pdata.frame
##  to find out which, use, e.g., table(index(your_pdataframe), useNA = "ifany")
```

```
summary(fixed_effects)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = repcoughing ~ educ + female + age + age2, data = woodstove,
##     model = "within", index = c("hh_id", "year"))
##
## Unbalanced Panel: n = 445, T = 1-18, N = 3045
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -0.948054 -0.220501 -0.059197  0.139354  0.966891
##
## Coefficients:
##           Estimate  Std. Error t-value  Pr(>|t|)
## educ   -8.5506e-03  3.1982e-03 -2.6736  0.007552 **
## female  4.0389e-02  1.5004e-02  2.6919  0.007151 **
## age    -1.8897e-03  1.4223e-03 -1.3286  0.184109
## age2    8.6117e-05  2.1982e-05  3.9176 9.173e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    415.47
## Residual Sum of Squares: 399.88
## R-Squared:      0.037526
## Adj. R-Squared: -0.12857
## F-statistic: 25.3042 on 4 and 2596 DF, p-value: < 2.22e-16
```

```
# run random effects (re)
random_effects <- plm(repcoughing ~ educ + female + age + age2, data = woodstove, model = "random", ind
```

```
## Warning in pdata.frame(data, index): duplicate couples (id-time) in resulting pdata.frame
##  to find out which, use, e.g., table(index(your_pdataframe), useNA = "ifany")
```

```
summary(random_effects)
```

```
## Oneway (individual) effect Random Effect Model
```

```
##    (Swamy-Arora's transformation)
##
## Call:
## plm(formula = repcoughing ~ educ + female + age + age2, data = woodstove,
##     model = "random", index = c("hh_id", "year"))
##
## Unbalanced Panel: n = 445, T = 1-18, N = 3045
##
## Effects:
##                   var std.dev share
## idiosyncratic 0.15404 0.39248 0.773
## individual    0.04514 0.21247 0.227
## theta:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1206  0.3979  0.4757  0.4502  0.5134  0.6008
##
## Residuals:
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -0.75113 -0.26887 -0.17796 -0.00364  0.41872  0.89615
##
## Coefficients:
##               Estimate  Std. Error z-value  Pr(>|z|)
## (Intercept)  2.6948e-01  2.2848e-02 11.7943 < 2.2e-16 ***
## educ        -9.3607e-03  2.9550e-03 -3.1677 0.0015365 **
## female       3.7510e-02  1.4858e-02  2.5246 0.0115829 *
## age         -9.5556e-04  1.3556e-03 -0.7049 0.4808699
## age2         7.4336e-05  2.0211e-05  3.6779 0.0002351 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    502.21
## Residual Sum of Squares: 474.26
## R-Squared:      0.056037
## Adj. R-Squared: 0.054795
## Chisq: 128.808 on 4 DF, p-value: < 2.22e-16
```

```
#Part c
hausman_test <- phtest(fixed_effects, random_effects)
print(hausman_test)
```

```
##
##  Hausman Test
##
## data:  repcoughing ~ educ + female + age + age2
## chisq = 10.43, df = 4, p-value = 0.03377
## alternative hypothesis: one model is inconsistent
```

```
#Question 3
#upload all the neccessary libraries
library(haven)
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
# Question 3.a)
# Load the dataset
woodstove <- read_dta("woodstove.dta")
# run the linear probability Model
lp_model <- lm(repcoughing ~ openfire + female + child + age, data = woodstove)
# Run the logit model
logit_model <- glm(repcoughing ~ openfire + female + child + age, data = woodstove, family = binomial(li
# Run the Poisson Model
poisson_model <- glm(repcoughing ~ openfire + female + child + age, data = woodstove, family = poisson()
# Display Poisson regression results
summary(poisson_model)
```

```
##
## Call:
## glm(formula = repcoughing ~ openfire + female + child + age,
##     family = poisson(), data = woodstove)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.575998   0.109515 -14.391  < 2e-16 ***
## openfire     0.158501   0.059038   2.685  0.00726 **
## female       0.096107   0.056074   1.714  0.08654 .
## child        0.096842   0.088728   1.091  0.27508
## age          0.009732   0.002272   4.283 1.84e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 2958.8  on 4066  degrees of freedom
## Residual deviance: 2920.4  on 4062  degrees of freedom
##   (118 observations deleted due to missingness)
## AIC: 5486.4
##
## Number of Fisher Scoring iterations: 5
```

```
#Question 3.b)
# Load required library
library(margins)
```

```
# Calculate marginal effects
marginal_effects <- margins(logit_model)
# Print the marginal effects
summary(marginal_effects)
```

```
##    factor    AME     SE      z      p   lower  upper
##       age 0.0033 0.0006 5.3175 0.0000  0.0021 0.0045
##     child 0.0353 0.0231 1.5279 0.1265 -0.0100 0.0806
```

```
##    female 0.0302 0.0144 2.0892 0.0367  0.0019 0.0585
##  openfire 0.0492 0.0150 3.2719 0.0011  0.0197 0.0786
```

```r
#Question 3.c)
# Perform Wald test
wald_test <- anova(logit_model, test = "Chisq")
print(wald_test)
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: repcoughing
##
## Terms added sequentially (first to last)
##
##
##          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                      4066     5063.0
## openfire  1  10.5491      4065     5052.4  0.001162 **
## female    1   3.5689      4064     5048.9  0.058870 .
## child     1  16.2136      4063     5032.7 5.659e-05 ***
## age       1  27.6159      4062     5005.0 1.480e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#Question 3.d)
# Load the required libraries
library(haven)
library(MASS)
library(margins)
library(lmtest)
# Load the dataset
woodstove <- read_dta("woodstove.dta")
# Create a subset of the data without missing values in openfire, female, child, and age
subset_data <- na.omit(woodstove[c("repcoughing", "openfire", "female", "child", "age")])
# Logit Model
logit_model <- glm(repcoughing ~ openfire + female + child + age, data = subset_data, family = binomial
# Perform likelihood ratio test
lr_test <- lrtest(logit_model, ~ openfire + female + child + age)
print(lr_test)
```

```
## Likelihood ratio test
##
## Model 1: repcoughing ~ openfire + female + child + age
## Model 2: repcoughing ~ openfire + female + child + age
##   #Df  LogLik Df Chisq Pr(>Chisq)
## 1   5 -2502.5
## 2   5 -2502.5  0     0          1
```

```r
#Question 3.e)
# Create a new variable backpain
```

```r
woodstove$backpain <- ifelse(woodstove$backpaindays > 0 & !is.na(woodstove$backpaindays), 1, 0)
# Replace age-squared (age2) with child
woodstove$child <- ifelse(woodstove$backpain == 1, 0, woodstove$child)
# Logit Model with backpain
logit_backpain_model <- glm(repcoughing ~ openfire + female + child + age, data = woodstove, family = bi
logit_backpain_model
```

```
##
## Call:  glm(formula = repcoughing ~ openfire + female + child + age,
##     family = binomial(link = "logit"), data = woodstove)
##
## Coefficients:
## (Intercept)     openfire       female        child          age
##    -1.42254      0.23152      0.14214      0.16627      0.01548
##
## Degrees of Freedom: 4066 Total (i.e. Null);  4062 Residual
##   (118 observations deleted due to missingness)
## Null Deviance:       5063
## Residual Deviance: 5005  AIC: 5015
```