Hypothesis\_Testing

#Previous Anlysis shown why the fixed effect model should be used #Also helped you to get to grips with various relevent packages and functions

#Reload of some relevent packages

library(tidyverse) # Modern data science library

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.4.2 ✔ purrr 0.3.4  
## ✔ tibble 3.2.1 ✔ dplyr 1.1.1  
## ✔ tidyr 1.2.0 ✔ stringr 1.4.0  
## ✔ readr 2.1.3 ✔ forcats 0.5.1  
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(plm) # Panel data analysis library

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

library(car) # Companion to applied regression

## 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

library(gplots) # Various programming tools for plotting data

##   
## Attaching package: 'gplots'  
##   
## The following object is masked from 'package:stats':  
##   
## lowess

library(tseries) # For timeseries analysis

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

library(lmtest) # For hetoroskedasticity analysis

## Loading required package: zoo  
##   
## Attaching package: 'zoo'  
##   
## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

library(openxlsx) # Load excel files  
library(writexl) # Write excel files  
library(stargazer) # Create nice tables

##   
## Please cite as:   
##   
## Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.  
## R package version 5.2.3. https://CRAN.R-project.org/package=stargazer

library(urca) #Unit root testing

#Load data

SS\_imputed\_data\_final\_logged <-read.xlsx("final\_sample\_imputed\_logged\_data.xlsx")

#Creating a full model

fixed\_time <- plm(CO2 ~ EU\_Exp + nonEU\_AnnexI\_Exp + GDP + Urban\_Pop + Renewables + Manufacturing + FDI, data= select(SS\_imputed\_data\_final\_logged, -c(.imp, .id)),index = c("Name", "Year"), model="within", effect = "time")  
summary(fixed\_time)

## Oneway (time) effect Within Model  
##   
## Call:  
## plm(formula = CO2 ~ EU\_Exp + nonEU\_AnnexI\_Exp + GDP + Urban\_Pop +   
## Renewables + Manufacturing + FDI, data = select(SS\_imputed\_data\_final\_logged,   
## -c(.imp, .id)), effect = "time", model = "within", index = c("Name",   
## "Year"))  
##   
## Unbalanced Panel: n = 116, T = 7-21, N = 1920  
##   
## Residuals:  
## Min. 1st Qu. Median 3rd Qu. Max.   
## -4.306252 -0.468979 -0.023718 0.452721 3.614316   
##   
## Coefficients:  
## Estimate Std. Error t-value Pr(>|t|)   
## EU\_Exp -0.030627 0.011956 -2.5616 0.01050 \*   
## nonEU\_AnnexI\_Exp -0.019076 0.013153 -1.4503 0.14715   
## GDP 0.933234 0.023494 39.7224 < 2.2e-16 \*\*\*  
## Urban\_Pop 0.070097 0.014894 4.7065 2.704e-06 \*\*\*  
## Renewables -0.565049 0.040011 -14.1223 < 2.2e-16 \*\*\*  
## Manufacturing 0.020966 0.010894 1.9245 0.05444 .   
## FDI 0.042912 0.028732 1.4935 0.13546   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Total Sum of Squares: 7707.8  
## Residual Sum of Squares: 1184.1  
## R-Squared: 0.84637  
## Adj. R-Squared: 0.84393  
## F-statistic: 1486.67 on 7 and 1889 DF, p-value: < 2.22e-16

fixed\_twoways <- plm(CO2 ~ EU\_Exp + nonEU\_AnnexI\_Exp + GDP + Urban\_Pop + Renewables + Manufacturing + FDI, data= select(SS\_imputed\_data\_final\_logged, -c(.imp, .id)),index = c("Name", "Year"), model="within", effect = "twoways")  
summary(fixed\_twoways)

## Twoways effects Within Model  
##   
## Call:  
## plm(formula = CO2 ~ EU\_Exp + nonEU\_AnnexI\_Exp + GDP + Urban\_Pop +   
## Renewables + Manufacturing + FDI, data = select(SS\_imputed\_data\_final\_logged,   
## -c(.imp, .id)), effect = "twoways", model = "within", index = c("Name",   
## "Year"))  
##   
## Unbalanced Panel: n = 116, T = 7-21, N = 1920  
##   
## Residuals:  
## Min. 1st Qu. Median 3rd Qu. Max.   
## -1.9610424 -0.0839716 0.0026691 0.0877160 1.0432810   
##   
## Coefficients:  
## Estimate Std. Error t-value Pr(>|t|)   
## EU\_Exp -0.0129489 0.0063646 -2.0345 0.042046 \*   
## nonEU\_AnnexI\_Exp -0.0111932 0.0039145 -2.8594 0.004294 \*\*   
## GDP 0.1430706 0.0160668 8.9047 < 2.2e-16 \*\*\*  
## Urban\_Pop 0.0043563 0.0073612 0.5918 0.554063   
## Renewables 0.1892528 0.0701591 2.6975 0.007053 \*\*   
## Manufacturing 0.0146355 0.0072339 2.0232 0.043204 \*   
## FDI 0.0058076 0.0122927 0.4724 0.636669   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Total Sum of Squares: 69.25  
## Residual Sum of Squares: 64.271  
## R-Squared: 0.071906  
## Adj. R-Squared: -0.0039526  
## F-statistic: 19.635 on 7 and 1774 DF, p-value: < 2.22e-16

# When running the two way effect, also considering country differences, very low R2 and more reason to doubt Brussels Effect since non\_EU\_Annex 1 more statistically significant, higher p value. However other such models seem problematic and didn't produce the signs we could expect e.g Renewable. The time effect model produced the right signs  
  
#strongest model which takes adjusted R2 form 83% to 84% is GDP, Urban Pop, Renewable

#Fitting our full fixed model

fixed\_time <- plm(CO2 ~ EU\_Exp + nonEU\_AnnexI\_Exp + GDP + Urban\_Pop + Renewables + Manufacturing + FDI, data= select(SS\_imputed\_data\_final\_logged, -c(.imp, .id)),index = c("Name", "Year"), model="within", effect = "time")  
summary(fixed\_time)

## Oneway (time) effect Within Model  
##   
## Call:  
## plm(formula = CO2 ~ EU\_Exp + nonEU\_AnnexI\_Exp + GDP + Urban\_Pop +   
## Renewables + Manufacturing + FDI, data = select(SS\_imputed\_data\_final\_logged,   
## -c(.imp, .id)), effect = "time", model = "within", index = c("Name",   
## "Year"))  
##   
## Unbalanced Panel: n = 116, T = 7-21, N = 1920  
##   
## Residuals:  
## Min. 1st Qu. Median 3rd Qu. Max.   
## -4.306252 -0.468979 -0.023718 0.452721 3.614316   
##   
## Coefficients:  
## Estimate Std. Error t-value Pr(>|t|)   
## EU\_Exp -0.030627 0.011956 -2.5616 0.01050 \*   
## nonEU\_AnnexI\_Exp -0.019076 0.013153 -1.4503 0.14715   
## GDP 0.933234 0.023494 39.7224 < 2.2e-16 \*\*\*  
## Urban\_Pop 0.070097 0.014894 4.7065 2.704e-06 \*\*\*  
## Renewables -0.565049 0.040011 -14.1223 < 2.2e-16 \*\*\*  
## Manufacturing 0.020966 0.010894 1.9245 0.05444 .   
## FDI 0.042912 0.028732 1.4935 0.13546   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Total Sum of Squares: 7707.8  
## Residual Sum of Squares: 1184.1  
## R-Squared: 0.84637  
## Adj. R-Squared: 0.84393  
## F-statistic: 1486.67 on 7 and 1889 DF, p-value: < 2.22e-16

#All the signs seem right, no indication of multicollinearity  
# # Interpret all esp EU\_Exp coefficient of -0.030627   
#Formally judge fit of full model fixed\_time using F test  
#We can formally judge the fit of the adjusted R2 value in our full\_model where the formal test is given by the result:  
# H0:R2=0 the set of explanatory variables are insignificant  
# H1:R2>0 the set of explanatory variables are significant  
#The F-test decision rule can be summarized as:  
# if F\_calc<F\_crit=>H\_0  
# if F\_calc>F\_crit=>H\_1  
  
#F stats: F Calc = 486.67, on 7 and 1889 DF  
  
fixed\_time\_qf <- qf(0.95, 7, 1889)  
print(fixed\_time\_qf) #F Critical= 2.014419, F cacl is larger than F stat

## [1] 2.014419

# 2.014419 < 486.67 -> H1 , implying that this is a valid model. Though this is a relatively weak test, since R2 only has to be greater than 0, the model is worthy of further investigation.  
  
  
#We should now begin to evaluate the importance of any one single explanatory variable. I can immediately see that nonEU\_AnnexI\_Exp and FDI are the statistical least significant coefficients in the table with pvalues of 0.14715 and 0.13546 respectively. Remove nonEU\_AnnexI\_Exp first since it is the largest p value  
  
#We will use a two tailed t-test since there should be and indeed there is a small negative relationship between nonEU\_AnnexI\_Exp and CO2. We would not necessarily expect this  
  
 #H0:b=0 explanatory variable is not important  
 #Ha:b≠0 explanatory variable has an influence  
 #if -t\_crit<t\_calc<+t\_crit=>H0  
 #if t\_calc<-t\_crit,or,t\_calc>+tcrit=>H1  
  
  
#tcalc for nonEU\_AnnexI\_Exp\_qt is -1.4503  
nonEU\_AnnexI\_Exp\_qt <- qt(0.95, 1889) # 1889 DF  
print(nonEU\_AnnexI\_Exp\_qt) # tcrit is 1.645661

## [1] 1.645661

#-tcrit(-1.645661) < (tcalc) -1.4503 < (+tcrit) 1.645661 -> H0  
  
  
#remove the variable  
  
  
#------  
  
# FDI t test, positive relationship withr response variable, right tailed t test  
  
 #H0:b1=0 explanatory variable is not important  
 #H1:b1>0 explanatory variable has a positive influence  
 #if t\_calc<t\_crit=>H0  
 #if t\_calc>t\_crit=>H1  
  
#tcalc= 1.4935  
FDI\_qt <-qt(0.95,1889 )  
print(FDI\_qt) # tcrit= 1.645661

## [1] 1.645661

#(tcrit)1.645661 > (tcalc)1.4935 therefore H1 and keep the variable  
  
  
#opt for lowest p value, want as simpler model as possible in spirit of parsimony

fixed\_time\_2 <- update(fixed\_time, . ~ . - nonEU\_AnnexI\_Exp)  
summary(fixed\_time\_2)

## Oneway (time) effect Within Model  
##   
## Call:  
## plm(formula = CO2 ~ EU\_Exp + GDP + Urban\_Pop + Renewables + Manufacturing +   
## FDI, data = select(SS\_imputed\_data\_final\_logged, -c(.imp,   
## .id)), effect = "time", model = "within", index = c("Name",   
## "Year"))  
##   
## Unbalanced Panel: n = 116, T = 7-21, N = 1920  
##   
## Residuals:  
## Min. 1st Qu. Median 3rd Qu. Max.   
## -4.247126 -0.476500 -0.020374 0.452040 3.606319   
##   
## Coefficients:  
## Estimate Std. Error t-value Pr(>|t|)   
## EU\_Exp -0.030987 0.011957 -2.5915 0.00963 \*\*   
## GDP 0.939945 0.023040 40.7955 < 2.2e-16 \*\*\*  
## Urban\_Pop 0.066287 0.014665 4.5202 6.559e-06 \*\*\*  
## Renewables -0.577388 0.039107 -14.7641 < 2.2e-16 \*\*\*  
## Manufacturing 0.021108 0.010897 1.9371 0.05289 .   
## FDI 0.048498 0.028481 1.7028 0.08877 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Total Sum of Squares: 7707.8  
## Residual Sum of Squares: 1185.5  
## R-Squared: 0.8462  
## Adj. R-Squared: 0.84384  
## F-statistic: 1733.09 on 6 and 1890 DF, p-value: < 2.22e-16

fixed\_time\_3 <- update(fixed\_time\_2, . ~ . - FDI)  
summary(fixed\_time\_3)

## Oneway (time) effect Within Model  
##   
## Call:  
## plm(formula = CO2 ~ EU\_Exp + GDP + Urban\_Pop + Renewables + Manufacturing,   
## data = select(SS\_imputed\_data\_final\_logged, -c(.imp, .id)),   
## effect = "time", model = "within", index = c("Name", "Year"))  
##   
## Unbalanced Panel: n = 116, T = 7-21, N = 1920  
##   
## Residuals:  
## Min. 1st Qu. Median 3rd Qu. Max.   
## -4.343304 -0.473165 -0.019649 0.459103 3.547510   
##   
## Coefficients:  
## Estimate Std. Error t-value Pr(>|t|)   
## EU\_Exp -0.029320 0.011923 -2.4591 0.01402 \*   
## GDP 0.944032 0.022927 41.1762 < 2.2e-16 \*\*\*  
## Urban\_Pop 0.067520 0.014654 4.6076 4.345e-06 \*\*\*  
## Renewables -0.580208 0.039092 -14.8421 < 2.2e-16 \*\*\*  
## Manufacturing 0.024695 0.010697 2.3087 0.02107 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Total Sum of Squares: 7707.8  
## Residual Sum of Squares: 1187.3  
## R-Squared: 0.84596  
## Adj. R-Squared: 0.84368  
## F-statistic: 2077.04 on 5 and 1891 DF, p-value: < 2.22e-16

fixed\_time\_4 <- update(fixed\_time\_3, . ~ . - Manufacturing)  
summary(fixed\_time\_4)

## Oneway (time) effect Within Model  
##   
## Call:  
## plm(formula = CO2 ~ EU\_Exp + GDP + Urban\_Pop + Renewables, data = select(SS\_imputed\_data\_final\_logged,   
## -c(.imp, .id)), effect = "time", model = "within", index = c("Name",   
## "Year"))  
##   
## Unbalanced Panel: n = 116, T = 7-21, N = 1920  
##   
## Residuals:  
## Min. 1st Qu. Median 3rd Qu. Max.   
## -4.366085 -0.450263 -0.020484 0.462949 3.537520   
##   
## Coefficients:  
## Estimate Std. Error t-value Pr(>|t|)   
## EU\_Exp -0.029230 0.011937 -2.4487 0.01443 \*   
## GDP 0.947087 0.022915 41.3312 < 2.2e-16 \*\*\*  
## Urban\_Pop 0.070146 0.014626 4.7958 1.747e-06 \*\*\*  
## Renewables -0.605274 0.037597 -16.0990 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Total Sum of Squares: 7707.8  
## Residual Sum of Squares: 1190.6  
## R-Squared: 0.84553  
## Adj. R-Squared: 0.84332  
## F-statistic: 2589.05 on 4 and 1892 DF, p-value: < 2.22e-16

# T test of EU\_Exp, negative relationship   
  
 #H0:b=0 explanatory variable is not important  
 #H1:b<0 explanatory variable has a negative influence  
 #if t\_calc>t\_crit=>H0  
 #if t\_calc<t\_crit=>H1  
  
#tcalc -2.4487  
  
EU\_Exp\_qt <- qt(0.95, 1892) # 1892 DF  
print(EU\_Exp\_qt) #tcrit= 1.645659

## [1] 1.645659

# -2.4487 < t\_crit -> H1 we keep the variable   
#the p vlaue is also significant at 0.95 level, acceptable to keep the variable in, even if it is not the mos important explanatory variable (see below)  
# Unlike nonEU\_AnnexI\_Exp which did not have a significant p value  
  
  
  
fixed\_time\_5 <- update(fixed\_time\_4, . ~ . - EU\_Exp)  
summary(fixed\_time\_5)

## Oneway (time) effect Within Model  
##   
## Call:  
## plm(formula = CO2 ~ GDP + Urban\_Pop + Renewables, data = select(SS\_imputed\_data\_final\_logged,   
## -c(.imp, .id)), effect = "time", model = "within", index = c("Name",   
## "Year"))  
##   
## Unbalanced Panel: n = 116, T = 7-21, N = 1920  
##   
## Residuals:  
## Min. 1st Qu. Median 3rd Qu. Max.   
## -4.307313 -0.452282 -0.018389 0.456017 3.468687   
##   
## Coefficients:  
## Estimate Std. Error t-value Pr(>|t|)   
## GDP 0.922154 0.020555 44.8624 < 2.2e-16 \*\*\*  
## Urban\_Pop 0.060554 0.014111 4.2913 1.866e-05 \*\*\*  
## Renewables -0.615241 0.037425 -16.4392 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Total Sum of Squares: 7707.8  
## Residual Sum of Squares: 1194.4  
## R-Squared: 0.84504  
## Adj. R-Squared: 0.84291  
## F-statistic: 3440.98 on 3 and 1893 DF, p-value: < 2.22e-16

fixed\_time\_6 <- update(fixed\_time\_5, . ~ . - Urban\_Pop)  
summary(fixed\_time\_6)

## Oneway (time) effect Within Model  
##   
## Call:  
## plm(formula = CO2 ~ GDP + Renewables, data = select(SS\_imputed\_data\_final\_logged,   
## -c(.imp, .id)), effect = "time", model = "within", index = c("Name",   
## "Year"))  
##   
## Unbalanced Panel: n = 116, T = 7-21, N = 1920  
##   
## Residuals:  
## Min. 1st Qu. Median 3rd Qu. Max.   
## -4.435861 -0.457273 -0.010493 0.461468 3.281071   
##   
## Coefficients:  
## Estimate Std. Error t-value Pr(>|t|)   
## GDP 0.999058 0.010114 98.780 < 2.2e-16 \*\*\*  
## Renewables -0.624828 0.037530 -16.649 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Total Sum of Squares: 7707.8  
## Residual Sum of Squares: 1206  
## R-Squared: 0.84353  
## Adj. R-Squared: 0.84147  
## F-statistic: 5105.32 on 2 and 1894 DF, p-value: < 2.22e-16

fixed\_time\_7 <- update(fixed\_time\_6, . ~ . - Renewables)  
summary(fixed\_time\_7)

## Oneway (time) effect Within Model  
##   
## Call:  
## plm(formula = CO2 ~ GDP, data = select(SS\_imputed\_data\_final\_logged,   
## -c(.imp, .id)), effect = "time", model = "within", index = c("Name",   
## "Year"))  
##   
## Unbalanced Panel: n = 116, T = 7-21, N = 1920  
##   
## Residuals:  
## Min. 1st Qu. Median 3rd Qu. Max.   
## -5.011533 -0.545030 -0.041372 0.535649 2.711620   
##   
## Coefficients:  
## Estimate Std. Error t-value Pr(>|t|)   
## GDP 0.936949 0.010063 93.112 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Total Sum of Squares: 7707.8  
## Residual Sum of Squares: 1382.5  
## R-Squared: 0.82063  
## Adj. R-Squared: 0.81836  
## F-statistic: 8669.88 on 1 and 1895 DF, p-value: < 2.22e-16

# Create a list of models  
models\_list <- list(fixed\_time, fixed\_time\_2, fixed\_time\_3, fixed\_time\_4, fixed\_time\_5, fixed\_time\_6, fixed\_time\_7)  
  
# Generate the table  
stargazer(models\_list, title="Fixed Effect Models", type="html", header=TRUE, out = "model\_graphic\_1.html")

##   
## <table style="text-align:center"><caption><strong>Fixed Effect Models</strong></caption>  
## <tr><td colspan="8" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left"></td><td colspan="7"><em>Dependent variable:</em></td></tr>  
## <tr><td></td><td colspan="7" style="border-bottom: 1px solid black"></td></tr>  
## <tr><td style="text-align:left"></td><td colspan="7">CO2</td></tr>  
## <tr><td style="text-align:left"></td><td>(1)</td><td>(2)</td><td>(3)</td><td>(4)</td><td>(5)</td><td>(6)</td><td>(7)</td></tr>  
## <tr><td colspan="8" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left">EU\_Exp</td><td>-0.031<sup>\*\*</sup></td><td>-0.031<sup>\*\*\*</sup></td><td>-0.029<sup>\*\*</sup></td><td>-0.029<sup>\*\*</sup></td><td></td><td></td><td></td></tr>  
## <tr><td style="text-align:left"></td><td>(0.012)</td><td>(0.012)</td><td>(0.012)</td><td>(0.012)</td><td></td><td></td><td></td></tr>  
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>  
## <tr><td style="text-align:left">nonEU\_AnnexI\_Exp</td><td>-0.019</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>  
## <tr><td style="text-align:left"></td><td>(0.013)</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>  
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>  
## <tr><td style="text-align:left">GDP</td><td>0.933<sup>\*\*\*</sup></td><td>0.940<sup>\*\*\*</sup></td><td>0.944<sup>\*\*\*</sup></td><td>0.947<sup>\*\*\*</sup></td><td>0.922<sup>\*\*\*</sup></td><td>0.999<sup>\*\*\*</sup></td><td>0.937<sup>\*\*\*</sup></td></tr>  
## <tr><td style="text-align:left"></td><td>(0.023)</td><td>(0.023)</td><td>(0.023)</td><td>(0.023)</td><td>(0.021)</td><td>(0.010)</td><td>(0.010)</td></tr>  
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>  
## <tr><td style="text-align:left">Urban\_Pop</td><td>0.070<sup>\*\*\*</sup></td><td>0.066<sup>\*\*\*</sup></td><td>0.068<sup>\*\*\*</sup></td><td>0.070<sup>\*\*\*</sup></td><td>0.061<sup>\*\*\*</sup></td><td></td><td></td></tr>  
## <tr><td style="text-align:left"></td><td>(0.015)</td><td>(0.015)</td><td>(0.015)</td><td>(0.015)</td><td>(0.014)</td><td></td><td></td></tr>  
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>  
## <tr><td style="text-align:left">Renewables</td><td>-0.565<sup>\*\*\*</sup></td><td>-0.577<sup>\*\*\*</sup></td><td>-0.580<sup>\*\*\*</sup></td><td>-0.605<sup>\*\*\*</sup></td><td>-0.615<sup>\*\*\*</sup></td><td>-0.625<sup>\*\*\*</sup></td><td></td></tr>  
## <tr><td style="text-align:left"></td><td>(0.040)</td><td>(0.039)</td><td>(0.039)</td><td>(0.038)</td><td>(0.037)</td><td>(0.038)</td><td></td></tr>  
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>  
## <tr><td style="text-align:left">Manufacturing</td><td>0.021<sup>\*</sup></td><td>0.021<sup>\*</sup></td><td>0.025<sup>\*\*</sup></td><td></td><td></td><td></td><td></td></tr>  
## <tr><td style="text-align:left"></td><td>(0.011)</td><td>(0.011)</td><td>(0.011)</td><td></td><td></td><td></td><td></td></tr>  
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>  
## <tr><td style="text-align:left">FDI</td><td>0.043</td><td>0.048<sup>\*</sup></td><td></td><td></td><td></td><td></td><td></td></tr>  
## <tr><td style="text-align:left"></td><td>(0.029)</td><td>(0.028)</td><td></td><td></td><td></td><td></td><td></td></tr>  
## <tr><td style="text-align:left"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>  
## <tr><td colspan="8" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left">Observations</td><td>1,920</td><td>1,920</td><td>1,920</td><td>1,920</td><td>1,920</td><td>1,920</td><td>1,920</td></tr>  
## <tr><td style="text-align:left">R<sup>2</sup></td><td>0.846</td><td>0.846</td><td>0.846</td><td>0.846</td><td>0.845</td><td>0.844</td><td>0.821</td></tr>  
## <tr><td style="text-align:left">Adjusted R<sup>2</sup></td><td>0.844</td><td>0.844</td><td>0.844</td><td>0.843</td><td>0.843</td><td>0.841</td><td>0.818</td></tr>  
## <tr><td style="text-align:left">F Statistic</td><td>1,486.674<sup>\*\*\*</sup> (df = 7; 1889)</td><td>1,733.091<sup>\*\*\*</sup> (df = 6; 1890)</td><td>2,077.043<sup>\*\*\*</sup> (df = 5; 1891)</td><td>2,589.046<sup>\*\*\*</sup> (df = 4; 1892)</td><td>3,440.980<sup>\*\*\*</sup> (df = 3; 1893)</td><td>5,105.319<sup>\*\*\*</sup> (df = 2; 1894)</td><td>8,669.884<sup>\*\*\*</sup> (df = 1; 1895)</td></tr>  
## <tr><td colspan="8" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left"><em>Note:</em></td><td colspan="7" style="text-align:right"><sup>\*</sup>p<0.1; <sup>\*\*</sup>p<0.05; <sup>\*\*\*</sup>p<0.01</td></tr>  
## </table>

#Test your hypothesis H3, EU\_Exp model for SO2 and PM10

model\_SO2 <- plm(SO2 ~ EU\_Exp + GDP + Urban\_Pop + Renewables, data= select(SS\_imputed\_data\_final\_logged, -c(.imp, .id)),index = c("Name", "Year"), model="within", effect = "time")  
summary(model\_SO2)

## Oneway (time) effect Within Model  
##   
## Call:  
## plm(formula = SO2 ~ EU\_Exp + GDP + Urban\_Pop + Renewables, data = select(SS\_imputed\_data\_final\_logged,   
## -c(.imp, .id)), effect = "time", model = "within", index = c("Name",   
## "Year"))  
##   
## Unbalanced Panel: n = 116, T = 7-21, N = 1920  
##   
## Residuals:  
## Min. 1st Qu. Median 3rd Qu. Max.   
## -3.68682775 -0.55850386 -0.00070278 0.56196007 3.53606911   
##   
## Coefficients:  
## Estimate Std. Error t-value Pr(>|t|)   
## EU\_Exp 0.0036326 0.0148778 0.2442 0.807132   
## GDP 0.9952210 0.0285604 34.8462 < 2.2e-16 \*\*\*  
## Urban\_Pop -0.0496988 0.0182302 -2.7262 0.006466 \*\*   
## Renewables -0.0370600 0.0468603 -0.7909 0.429124   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Total Sum of Squares: 8141.9  
## Residual Sum of Squares: 1849.6  
## R-Squared: 0.77283  
## Adj. R-Squared: 0.76958  
## F-statistic: 1609.1 on 4 and 1892 DF, p-value: < 2.22e-16

model\_PM10 <- plm(PM10 ~ EU\_Exp + GDP + Urban\_Pop + Renewables, data= select(SS\_imputed\_data\_final\_logged, -c(.imp, .id)),index = c("Name", "Year"), model="within", effect = "time")  
summary(model\_PM10)

## Oneway (time) effect Within Model  
##   
## Call:  
## plm(formula = PM10 ~ EU\_Exp + GDP + Urban\_Pop + Renewables, data = select(SS\_imputed\_data\_final\_logged,   
## -c(.imp, .id)), effect = "time", model = "within", index = c("Name",   
## "Year"))  
##   
## Unbalanced Panel: n = 116, T = 7-21, N = 1920  
##   
## Residuals:  
## Min. 1st Qu. Median 3rd Qu. Max.   
## -6.019290 -0.900074 0.017262 0.945214 3.599104   
##   
## Coefficients:  
## Estimate Std. Error t-value Pr(>|t|)   
## EU\_Exp -0.246348 0.021415 -11.503 < 2.2e-16 \*\*\*  
## GDP 1.120423 0.041111 27.254 < 2.2e-16 \*\*\*  
## Urban\_Pop 0.189040 0.026241 7.204 8.407e-13 \*\*\*  
## Renewables -1.904934 0.067452 -28.241 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Total Sum of Squares: 11142  
## Residual Sum of Squares: 3832.3  
## R-Squared: 0.65604  
## Adj. R-Squared: 0.65113  
## F-statistic: 902.141 on 4 and 1892 DF, p-value: < 2.22e-16

summary(fixed\_time\_4)

## Oneway (time) effect Within Model  
##   
## Call:  
## plm(formula = CO2 ~ EU\_Exp + GDP + Urban\_Pop + Renewables, data = select(SS\_imputed\_data\_final\_logged,   
## -c(.imp, .id)), effect = "time", model = "within", index = c("Name",   
## "Year"))  
##   
## Unbalanced Panel: n = 116, T = 7-21, N = 1920  
##   
## Residuals:  
## Min. 1st Qu. Median 3rd Qu. Max.   
## -4.366085 -0.450263 -0.020484 0.462949 3.537520   
##   
## Coefficients:  
## Estimate Std. Error t-value Pr(>|t|)   
## EU\_Exp -0.029230 0.011937 -2.4487 0.01443 \*   
## GDP 0.947087 0.022915 41.3312 < 2.2e-16 \*\*\*  
## Urban\_Pop 0.070146 0.014626 4.7958 1.747e-06 \*\*\*  
## Renewables -0.605274 0.037597 -16.0990 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Total Sum of Squares: 7707.8  
## Residual Sum of Squares: 1190.6  
## R-Squared: 0.84553  
## Adj. R-Squared: 0.84332  
## F-statistic: 2589.05 on 4 and 1892 DF, p-value: < 2.22e-16

models\_list\_2 <- list(model\_SO2, model\_PM10)  
stargazer(models\_list\_2, title="Fixed Effect Models", type="html", header=TRUE, out = "model\_graphic\_2.html")

##   
## <table style="text-align:center"><caption><strong>Fixed Effect Models</strong></caption>  
## <tr><td colspan="3" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left"></td><td colspan="2"><em>Dependent variable:</em></td></tr>  
## <tr><td></td><td colspan="2" style="border-bottom: 1px solid black"></td></tr>  
## <tr><td style="text-align:left"></td><td>SO2</td><td>PM10</td></tr>  
## <tr><td style="text-align:left"></td><td>(1)</td><td>(2)</td></tr>  
## <tr><td colspan="3" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left">EU\_Exp</td><td>0.004</td><td>-0.246<sup>\*\*\*</sup></td></tr>  
## <tr><td style="text-align:left"></td><td>(0.015)</td><td>(0.021)</td></tr>  
## <tr><td style="text-align:left"></td><td></td><td></td></tr>  
## <tr><td style="text-align:left">GDP</td><td>0.995<sup>\*\*\*</sup></td><td>1.120<sup>\*\*\*</sup></td></tr>  
## <tr><td style="text-align:left"></td><td>(0.029)</td><td>(0.041)</td></tr>  
## <tr><td style="text-align:left"></td><td></td><td></td></tr>  
## <tr><td style="text-align:left">Urban\_Pop</td><td>-0.050<sup>\*\*\*</sup></td><td>0.189<sup>\*\*\*</sup></td></tr>  
## <tr><td style="text-align:left"></td><td>(0.018)</td><td>(0.026)</td></tr>  
## <tr><td style="text-align:left"></td><td></td><td></td></tr>  
## <tr><td style="text-align:left">Renewables</td><td>-0.037</td><td>-1.905<sup>\*\*\*</sup></td></tr>  
## <tr><td style="text-align:left"></td><td>(0.047)</td><td>(0.067)</td></tr>  
## <tr><td style="text-align:left"></td><td></td><td></td></tr>  
## <tr><td colspan="3" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left">Observations</td><td>1,920</td><td>1,920</td></tr>  
## <tr><td style="text-align:left">R<sup>2</sup></td><td>0.773</td><td>0.656</td></tr>  
## <tr><td style="text-align:left">Adjusted R<sup>2</sup></td><td>0.770</td><td>0.651</td></tr>  
## <tr><td style="text-align:left">F Statistic (df = 4; 1892)</td><td>1,609.104<sup>\*\*\*</sup></td><td>902.141<sup>\*\*\*</sup></td></tr>  
## <tr><td colspan="3" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left"><em>Note:</em></td><td colspan="2" style="text-align:right"><sup>\*</sup>p<0.1; <sup>\*\*</sup>p<0.05; <sup>\*\*\*</sup>p<0.01</td></tr>  
## </table>

#Check the assumptions of your model, fixed\_time\_4, and Regression Diagnostics from princeton

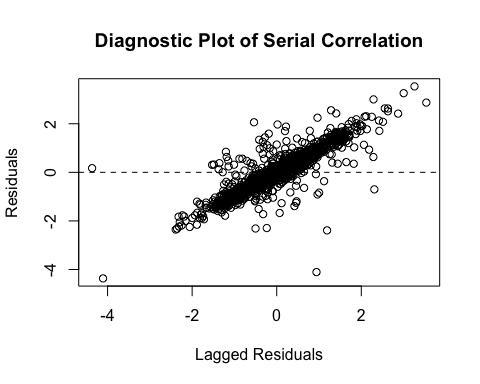
# Serial correlation ( Serial correlation tests apply to macro panels with long time series)  
  
# H0) The null is that there is not serial correlation  
  
pbgtest(fixed\_time\_4)

##   
## Breusch-Godfrey/Wooldridge test for serial correlation in panel models  
##   
## data: CO2 ~ EU\_Exp + GDP + Urban\_Pop + Renewables  
## chisq = 1517, df = 7, p-value < 2.2e-16  
## alternative hypothesis: serial correlation in idiosyncratic errors

#because p-value > 0.05, we conclude that there is NO serial correlation  
  
  
# Attempting to create a plot of serial correlation   
  
serial\_corr\_test <- plmtest(fixed\_time\_4, type = "bp", effect = "time")  
print(serial\_corr\_test)

##   
## Lagrange Multiplier Test - time effects (Breusch-Pagan)  
##   
## data: CO2 ~ EU\_Exp + GDP + Urban\_Pop + Renewables  
## chisq = 922.74, df = 1, p-value < 2.2e-16  
## alternative hypothesis: significant effects

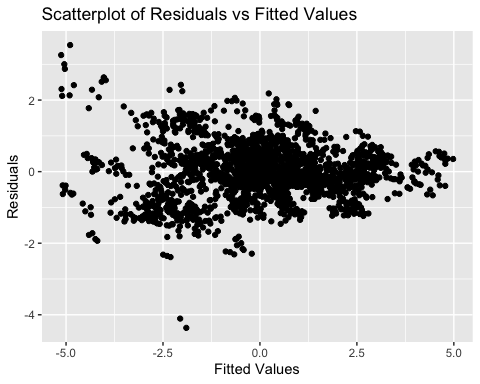
resid <- residuals(fixed\_time\_4)  
lagged\_resid <- c(rep(NA, 1), resid[-length(resid)])  
plot(lagged\_resid, resid, xlab = "Lagged Residuals", ylab = "Residuals", main = "Diagnostic Plot of Serial Correlation")  
abline(h = 0, lty = 2)



# Test for Heteroscedasticity in Panel Data Model  
  
# The null hypothesis for the Breusch-Pagan test is homoskedasticity.  
# If the p-value is less than 0.05, we reject the null hypothesis and conclude that   
# heteroscedasticity is present.  
  
het\_test <- bptest(fixed\_time\_4) # perform Breusch-Pagan test for heteroscedasticity  
print(het\_test)

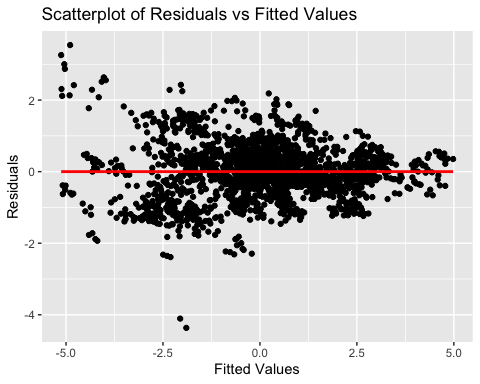
##   
## studentized Breusch-Pagan test  
##   
## data: fixed\_time\_4  
## BP = 295.43, df = 4, p-value < 2.2e-16

# Extract residuals and fitted values  
resid <- residuals(fixed\_time\_4)  
fitted <- fitted(fixed\_time\_4)  
  
# Create scatter plot using ggplot2  
ggplot(data.frame(fitted, resid), aes(x = fitted, y = resid)) +  
 geom\_point() +  
 labs(x = "Fitted Values", y = "Residuals") +  
 ggtitle("Scatterplot of Residuals vs Fitted Values")



# Add line of best fit to scatter plot  
ggplot(data.frame(fitted, resid), aes(x = fitted, y = resid)) +  
 geom\_point() +  
 geom\_smooth(method = "lm", se = FALSE, color = "red") +  
 labs(x = "Fitted Values", y = "Residuals") +  
 ggtitle("Scatterplot of Residuals vs Fitted Values")

## `geom\_smooth()` using formula = 'y ~ x'



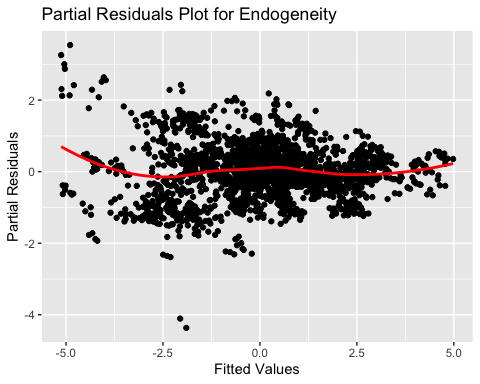
#In the plot, we are looking for a pattern in the distribution of residuals. Ideally, the residuals should be randomly distributed around the horizontal line at y = 0, indicating that there is no systematic pattern in the residuals. If there is a systematic pattern, such as a curve or a U-shape, this suggests that the model may not be a good fit for the data and that there may be some other variable that should be included in the model.  
  
#In addition, we can also check for heteroscedasticity in this plot by examining whether the spread of residuals is the same across the range of fitted values. If the spread of residuals increases or decreases as the fitted values increase, this indicates heteroscedasticity, which means that the variance of the residuals is not constant across the range of the independent variable(s).

#Endogeneity  
  
  
#is the test of Endogeneity and serial correlation the same/ very similar?  
  
endog\_test <- pbgtest(fixed\_time\_4, order = 1) #Arellano-Bond test for endogeneity  
print(endog\_test)

##   
## Breusch-Godfrey/Wooldridge test for serial correlation in panel models  
##   
## data: CO2 ~ EU\_Exp + GDP + Urban\_Pop + Renewables  
## chisq = 1508.6, df = 1, p-value < 2.2e-16  
## alternative hypothesis: serial correlation in idiosyncratic errors

#The Arellano-Bond test for endogeneity tests the null hypothesis that the variables included in the model are exogenous. A rejection of the null hypothesis indicates that at least one of the variables is endogenous and should be instrumented.  
  
#The test provides a p-value, and if the p-value is less than the chosen significance level (typically 0.05), then the null hypothesis is rejected and endogeneity is present in the model.  
  
#In addition, the test provides a test statistic, which can be used to assess the strength of the endogeneity. A higher test statistic suggests stronger evidence of endogeneity  
  
  
  
  
  
  
  
# Extract residuals and fitted values  
resid <- residuals(fixed\_time\_4)  
fitted <- fitted(fixed\_time\_4)  
  
# Create partial residuals plot using ggplot2  
ggplot(data.frame(fitted, resid), aes(x = fitted, y = resid)) +  
 geom\_point() +  
 geom\_smooth(method = "loess", se = FALSE, color = "red") +  
 labs(x = "Fitted Values", y = "Partial Residuals") +  
 ggtitle("Partial Residuals Plot for Endogeneity")

## `geom\_smooth()` using formula = 'y ~ x'



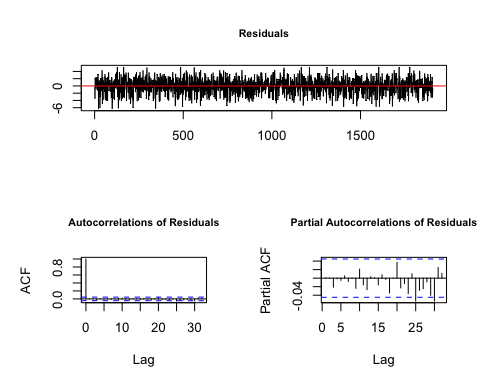
#In the scatterplot of residuals vs. fitted values for endogeneity, the horizontal line represents the relationship between the partial residuals and the fitted values after controlling for the other covariates in the model. If the line is flat, it suggests that there is no systematic relationship between the partial residuals and the fitted values, which is consistent with the assumption of exogeneity.  
  
#In the scatterplot of residuals vs. fitted values for heteroscedasticity, the horizontal line represents the mean of the residuals. If the line is straight, it suggests that the variance of the residuals is constant across the range of the fitted values, which is consistent with the assumption of homoscedasticity.

#Unit roots/stationarity testing  
  
#The Dickey-Fuller test to check for stochastic trends.  
  
#H0) The null hypothesis is that the series has a unit root (i.e. non-stationary)  
  
#If unit root is present you can take the first difference of the variable.  
  
adf.test(SS\_imputed\_data\_final\_logged$CO2, k=2)

## Warning in adf.test(SS\_imputed\_data\_final\_logged$CO2, k = 2): p-value smaller  
## than printed p-value

##   
## Augmented Dickey-Fuller Test  
##   
## data: SS\_imputed\_data\_final\_logged$CO2  
## Dickey-Fuller = -25.268, Lag order = 2, p-value = 0.01  
## alternative hypothesis: stationary

#Because p-value < 0.05, we conclude that the series does NOT have unit root. In other words, the series is stationary  
  
  
  
# Perform ADF test on CO2 variable in SS\_imputed\_data\_final\_logged data frame  
adf\_result <- ur.df(SS\_imputed\_data\_final\_logged$CO2, type = "trend", selectlags = "AIC", lags = 10)  
  
par(mar = c(5, 5, 4, 2) + 0.1, mfrow = c(3,1), cex.main = 0.8)  
  
# Plot ADF test results  
plot(adf\_result)



#Overall, a rapid decay of the ACF and a relatively small PACF can be taken as evidence that the residuals are approximately white noise, and that the model is doing a good job of accounting for the structure in the data.