Advanced Econometrics Project

Mattia Elezi

05/05/2022

Contents

Introduction

Panel data controls for the same entities over time throuh a time series dimension and a cross-sectional dimension.

```
library(lmtest)
library(texreg)
library(tidyr)
library(dplyr)
library(pdfetch)
library(foreign)
library(car)
library(gplots)
library(tseries)
library(sjPlot)
library(huxtable)
library(ivreg)
library(plm)

rm(list=ls())
```

Dataset

The variables available in the WDI file are the following: Country (total of 192 countries), Year (from 1960 to 2015), Income Per Capita (constant 2010 US), $GDPgrowth(annual)$, Oil Rents (oil rents as % of GDP), Safe Water Access (% of the population with access to safe water), Nurses Midwives (nurses and midwives per 1,000 people), Pregnant Women With Anemia (pregnant women with anemia in %), Maternal Mortality (maternal mortality ratio, modelled estimate, per 100,000 live births), Health Expenditure (total health expenditure as % of GDP), Poverty Gap (poverty gap at $1.90 a day, 2011 PPP, in %), and Infant Mortality (mortality rate, infant, per 1,000 live births).

```
wdi<-read.csv("C:/Users/cex/OneDrive/Desktop/panel data/wdi.csv",
na.strings = "NA")
```

Pooled OLS Model

The first model to be computed will be a panel model that ignores the panel effects, the data is going to be stacked and computed through OLS. The Pooled OLS model does not take into account the panel heterogeneity, ignoring the heterogeneity effect across countries, the panel effect across countries, and the fixed effect across countries. The result will be compared with the Fixed Effects (FE) and Random Effects (RE) models. The dependent variable is maternal mortality, the independent are the variables on the right-hand side of the regression equation.

```
pooled_OLS = plm(MaternalMortality ~ SafeWaterAccess + HealthExpenditure +
             PregnantWomenWithAnemia + IncomePerCapita,
                     data   = wdi,
                     index  = c("Country", "Year"),
                     model  = "pooling")


summary(pooled_OLS)
```

```
## Pooling Model
##
## Call:
## plm(formula = MaternalMortality ~ SafeWaterAccess + HealthExpenditure +
##     PregnantWomenWithAnemia + IncomePerCapita, data = wdi, model = "pooling",
##     index = c("Country", "Year"))
##
## Unbalanced Panel: n = 161, T = 1-17, N = 2076
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -541.1501  -72.2114   -3.2121   49.7978 1791.4759
##
## Coefficients:
##                            Estimate  Std. Error  t-value Pr(>|t|)
## (Intercept)              5.9534e+02  5.0486e+01  11.7921   <2e-16 ***
## SafeWaterAccess         -1.1344e+01  3.7973e-01 -29.8743   <2e-16 ***
## HealthExpenditure        2.9464e+01  2.1764e+00  13.5376   <2e-16 ***
## PregnantWomenWithAnemia  1.1644e+01  5.5042e-01  21.1553   <2e-16 ***
## IncomePerCapita         -8.3848e-05  4.7592e-04  -0.1762   0.8602
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    249470000
## Residual Sum of Squares: 77869000
## R-Squared:      0.68786
## Adj. R-Squared: 0.68726
## F-statistic: 1140.97 on 4 and 2071 DF, p-value: < 2.22e-16
```

The panel is unbalanced due to many missing values. Safe water access: the higher the percentage of the population of the country that accesses safe water, the better the infrastructure leading to lower maternal mortality. The hypothesis is confirmed by the results. Indeed, as can be inferred from the results, the coefficient is negative. When the country has a population with more access to safe water, maternal mortality decreases at the 0.1% significance level. Health expenditure: the higher the total health expenditure as % of the GDP of a country, the lower maternal mortality. The hypothesis is rejected by the results: the coefficient is positive, and the direct relationship is counterintuitive as when the country spends more of its GDP in the health sector, the maternal mortality ratio increases at the 0.1% significance level. The reason could be an endogeneity problem or the model could be not correctly specified because fixed and panelled effects are being ignored. An instrumental variable might be required to instrumentalise the variable because it might be endogenous. Pregnant women with anemia: The higher the percentage of women with anemia, the higher the index of maternal mortality. The hypothesis is confirmed by the results as when a larger percentage of women have anemia, the country experiences an increase in the maternal mortality index at the 0.1% significance level. Income per capita: the higher the GDP per capita, the lower the maternal mortality ratio. The sign of the result reflects the hypothesis, but the estimate is really small and has no significance.

Fixed Effects Models

Individual Fixed Effects Model

The dataset includes several indicators that could be related to maternal mortality rates such as healthcare expenditure, the proportion of the population with access to safe water, etc. However, a number of other factors could also have a profound effect on maternal mortality rates but they are either not easily available or not observable at all. Some of these factors could include a lack of awareness of preventable diseases and attitudes towards prevention or prenatal care that vary greatly from country to country but don't change over time within the same country. Without access to datasets controlling for these factors, the model would suffer from omitted variable bias. A fixed Effects (FE) model allows to control for these variables that are constant over time. The following model controls only for the within individual effect, not the time effect.

```
fixed_effects = plm(MaternalMortality ~ SafeWaterAccess + HealthExpenditure +
                    PregnantWomenWithAnemia + IncomePerCapita,
                     data   = wdi,
                     index  = c("Country", "Year"),
                     model  = "within",
                     effect = "individual")


summary(fixed_effects)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = MaternalMortality ~ SafeWaterAccess + HealthExpenditure +
##     PregnantWomenWithAnemia + IncomePerCapita, data = wdi, effect = "individual",
##     model = "within", index = c("Country", "Year"))
##
## Unbalanced Panel: n = 161, T = 1-17, N = 2076
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -564.5439  -12.9162    0.1942   11.6569  535.8183
##
## Coefficients:
##                            Estimate  Std. Error  t-value  Pr(>|t|)
```

```
## SafeWaterAccess         -8.39565735  0.54999320 -15.2650 < 2.2e-16 ***
## HealthExpenditure       -7.13888589  1.53446455  -4.6524 3.507e-06 ***
## PregnantWomenWithAnemia  6.08637540  0.78843707   7.7195 1.869e-14 ***
## IncomePerCapita          0.00325803  0.00097414   3.3445 0.0008402 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    10549000
## Residual Sum of Squares: 7206900
## R-Squared:      0.31681
## Adj. R-Squared: 0.25818
## F-statistic: 221.541 on 4 and 1911 DF, p-value: < 2.22e-16
```

Fixed effects consider the unobserved effects, and the model is partially controlling for the endogeneity problems using only individual effects. The direct relationship between health expenditure and maternal mortality has become negative and it is significant at the 0.1% level. The direct relationship between income per capita and maternal mortality has become positive and it is significant at the 0.1% level. Nevertheless, the estimate is really small. The adjusted R-squared is 0.25818.

Time Fixed Effects Model

In addition to country FE, several factors could affect maternal mortality rates that are not specific to each country. For example, the development of new vaccines or medicines to fight infections, or investments in awareness campaigns by aid organizations such as WHO about preventable diseases likely vary over time but have similar effects around the world. The following model controls only for the within time effect.

```r
time_effects = plm(MaternalMortality ~ SafeWaterAccess + HealthExpenditure +
                   PregnantWomenWithAnemia + IncomePerCapita,
                     data = wdi,
                     index = c("Country", "Year"),
                     model = "within",
                     effect = "time")

summary(time_effects)
```

```
## Oneway (time) effect Within Model
##
## Call:
## plm(formula = MaternalMortality ~ SafeWaterAccess + HealthExpenditure +
##     PregnantWomenWithAnemia + IncomePerCapita, data = wdi, effect = "time",
##     model = "within", index = c("Country", "Year"))
##
## Unbalanced Panel: n = 161, T = 1-17, N = 2076
##
## Residuals:
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -554.429  -71.522   -3.712   48.185 1806.502
##
## Coefficients:
##                            Estimate  Std. Error  t-value Pr(>|t|)
## SafeWaterAccess         -1.1353e+01  3.8119e-01 -29.7842   <2e-16 ***
## HealthExpenditure        2.9439e+01  2.2024e+00  13.3672   <2e-16 ***
```

```
## PregnantWomenWithAnemia  1.1662e+01  5.5271e-01  21.0996   <2e-16 ***
## IncomePerCapita         -8.5546e-05  4.7885e-04  -0.1786   0.8582
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    247430000
## Residual Sum of Squares: 77689000
## R-Squared:      0.68602
## Adj. R-Squared: 0.68297
## F-statistic: 1122.5 on 4 and 2055 DF, p-value: < 2.22e-16
```

The results reflect the results of the pooled OLS model. The adjusted R-squared is 0.68297. nevertheless, the Time Fixed Effects model does not distinguish between observations across different countries. And the betas present unexpected signs. When country fixed effects are not used, the fluctuations in the series over time make it seem as if an increase in health expenditure is correlated with an increase in maternal mortality, but in fact, it is just an artifact of time fixed effects.

The next step will be running a model that accounts for both country and time fixed effects.

Two-way Fixed Effects Model

```
twoway_effects = plm(MaternalMortality ~ SafeWaterAccess + HealthExpenditure
                     + PregnantWomenWithAnemia + IncomePerCapita,
                       data = wdi,
                       index = c("Country", "Year"),
                       model = "within",
                       effect = "twoways")

summary(twoway_effects)
```

```
## Twoways effects Within Model
##
## Call:
## plm(formula = MaternalMortality ~ SafeWaterAccess + HealthExpenditure +
##     PregnantWomenWithAnemia + IncomePerCapita, data = wdi, effect = "twoways",
##     model = "within", index = c("Country", "Year"))
##
## Unbalanced Panel: n = 161, T = 1-17, N = 2076
##
## Residuals:
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -567.582  -14.088    0.000   12.432  539.814
##
## Coefficients:
##                           Estimate Std. Error  t-value  Pr(>|t|)
## SafeWaterAccess         -7.9821605  0.6187568 -12.9003 < 2.2e-16 ***
## HealthExpenditure       -6.8293578  1.6087862  -4.2450 2.291e-05 ***
## PregnantWomenWithAnemia  5.5100821  0.9245709   5.9596 3.008e-09 ***
## IncomePerCapita          0.0042492  0.0011730   3.6225 0.0002994 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Total Sum of Squares:    8589400
## Residual Sum of Squares: 7173800
## R-Squared:      0.16481
## Adj. R-Squared: 0.085483
## F-statistic: 93.4894 on 4 and 1895 DF, p-value: < 2.22e-16
```

Once both individual (country) and time FE are included in the model, the results apart from income per capita which is positive, have the expected signs.

First Table of Comparison

Subsequently, a consolidated table comparing the three FE models and the Pooled OLS model for comparison side-by-side is created.

```
     screenreg(list(    pooled_OLS,
                        fixed_effects,
                        time_effects,
                        twoway_effects),

          custom.model.names = c("Pooled OLS",
                                 "Country FE",
                                 "Time FE",
                                 "Two-way FE"))
```

```
##
## ============================================================================
##                            Pooled OLS   Country FE   Time FE      Two-way FE
## ----------------------------------------------------------------------------
## (Intercept)                 595.34 ***
##                             (50.49)
## SafeWaterAccess             -11.34 ***    -8.40 ***   -11.35 ***    -7.98 ***
##                              (0.38)       (0.55)       (0.38)       (0.62)
## HealthExpenditure            29.46 ***    -7.14 ***    29.44 ***    -6.83 ***
##                              (2.18)       (1.53)       (2.20)       (1.61)
## PregnantWomenWithAnemia      11.64 ***     6.09 ***    11.66 ***     5.51 ***
##                              (0.55)       (0.79)       (0.55)       (0.92)
## IncomePerCapita              -0.00         0.00 ***    -0.00         0.00 ***
##                              (0.00)       (0.00)       (0.00)       (0.00)
## ----------------------------------------------------------------------------
## R^2                           0.69         0.32         0.69         0.16
## Adj. R^2                      0.69         0.26         0.68         0.09
## Num. obs.                  2076         2076         2076         2076
## ============================================================================
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

Every explanatory variable is statistically significant in each model, apart from income per capita. The coefficients for the explanatory variable in the two-way fixed effect model are close to the country fixed effects indicating that these factors vary greatly across countries more than they do across time.

Two-way Random Effects Model

In the RE model, the country-specific components are random and uncorrelated with the regressors. In this case, the “Two-way RE” model is computed, with both individual and time effects.

```
RE_twoway = plm( MaternalMortality ~ SafeWaterAccess + HealthExpenditure
                 + PregnantWomenWithAnemia + IncomePerCapita,

                        data   = wdi,
                        index  = c("Country", "Year"),
                        model  = "random",
                        effect = "twoways",
                 random.method = "swar" )


summary(RE_twoway)
```

```
## Twoways effects Random Effect Model
##    (Swamy-Arora's transformation)
##
## Call:
## plm(formula = MaternalMortality ~ SafeWaterAccess + HealthExpenditure +
##     PregnantWomenWithAnemia + IncomePerCapita, data = wdi, effect = "twoways",
##     model = "random", random.method = "swar", index = c("Country",
##         "Year"))
##
## Unbalanced Panel: n = 161, T = 1-17, N = 2076
##
## Effects:
##                      var   std.dev share
## idiosyncratic  3785.631    61.527 0.101
## individual    33612.758   183.338 0.898
## time             22.014     4.692 0.001
## theta:
##            Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## id    0.6818421 0.9188742 0.9188742 0.9121699 0.9188742 0.9188742
## time  0.2106073 0.2218063 0.2377456 0.2365045 0.2428450 0.2813444
## total 0.2103002 0.2216330 0.2374240 0.2360903 0.2425130 0.2810956
##
## Residuals:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -416.54  -77.74  -33.25    6.07    2.76 2169.25
##
## Coefficients:
##                           Estimate  Std. Error  z-value  Pr(>|z|)
## (Intercept)             7.5705e+02  1.0510e+00   720.32 < 2.2e-16 ***
## SafeWaterAccess        -8.9559e+00  8.1320e-03 -1101.32 < 2.2e-16 ***
## HealthExpenditure      -5.2444e+00  2.4669e-02  -212.59 < 2.2e-16 ***
## PregnantWomenWithAnemia  7.0088e+00  1.1942e-02   586.91 < 2.2e-16 ***
## IncomePerCapita         1.9027e-03  1.3284e-05   143.23 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Total Sum of Squares:    249470000
## Residual Sum of Squares: 97131000
## R-Squared:      0.64563
## Adj. R-Squared: 0.64495
## Chisq: 3796030 on 4 DF, p-value: < 2.22e-16
```

The RE model is computed by using the same function as before, but the effects are random and the country-fixed and time-fixed effects are included in the model. The panel is unbalanced. The signs of the coefficients, apart from income per capita, reflect the hypotheses stated at the beginning at the 0.1% significance level. The adjusted R-squared is 0.64495.

Second Table of Comparison

The following step will be the creation of a table to compare the Pooled OLS model, the FE Two-way model, and the RE Two-way model.

```
    screenreg(list(   pooled_OLS,
                      twoway_effects,
                      RE_twoway),

         custom.model.names = c("Pooled OLS",
                                "Two-way FE",
                                "Two-way RE" ))
```

```
##
## ================================================================
##                            Pooled OLS   Two-way FE   Two-way RE
## ----------------------------------------------------------------
## (Intercept)                 595.34 ***                757.05 ***
##                             (50.49)                    (1.05)
## SafeWaterAccess             -11.34 ***    -7.98 ***    -8.96 ***
##                              (0.38)       (0.62)       (0.01)
## HealthExpenditure            29.46 ***    -6.83 ***    -5.24 ***
##                              (2.18)       (1.61)       (0.02)
## PregnantWomenWithAnemia      11.64 ***     5.51 ***     7.01 ***
##                              (0.55)       (0.92)       (0.01)
## IncomePerCapita              -0.00         0.00 ***     0.00 ***
##                              (0.00)       (0.00)       (0.00)
## ----------------------------------------------------------------
## R^2                           0.69         0.16         0.65
## Adj. R^2                      0.69         0.09         0.64
## Num. obs.                  2076         2076         2076
## s_idios                                                61.53
## s_id                                                  183.34
## s_time                                                  4.69
## ================================================================
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

Lagged Dependent Variables (LDV) and Dynamic Models

The model addresses auto-correlation (AR) by modelling the time dependence directly. A dynamic model takes into account whether changes in the predictor variables have an immediate effect on our dependent variable or whether the effects are distributed over time. The value of the variable in the present is used to partially explain the behaviour of the variable in the past through the lagged variable. As panel data has two dimensions, cross-sectional and time series, the autoregressive nature of the variable should be controlled to efficiently analyse the properties of the regression.

```
ldv_model = plm(MaternalMortality ~ lag(MaternalMortality) + SafeWaterAccess
                + HealthExpenditure + PregnantWomenWithAnemia
                + IncomePerCapita,

                       data = wdi,
                      index = c("Country", "Year"),
                      model = "within",
                     effect = "twoways" )

summary(ldv_model)
```

```
## Twoways effects Within Model
##
## Call:
## plm(formula = MaternalMortality ~ lag(MaternalMortality) + SafeWaterAccess +
##     HealthExpenditure + PregnantWomenWithAnemia + IncomePerCapita,
##     data = wdi, effect = "twoways", model = "within", index = c("Country",
##         "Year"))
##
## Unbalanced Panel: n = 161, T = 1-17, N = 2076
##
## Residuals:
##        Min.     1st Qu.      Median     3rd Qu.        Max.
## -8.0916e+01 -1.7451e+00  3.1817e-12  2.0932e+00  8.9342e+01
##
## Coefficients:
##                            Estimate  Std. Error  t-value  Pr(>|t|)
## lag(MaternalMortality)   0.97518878  0.00357320 272.9175 < 2.2e-16 ***
## SafeWaterAccess         -0.18781989  0.10156134  -1.8493 0.0645667 .
## HealthExpenditure       -0.37611749  0.25450797  -1.4778 0.1396217
## PregnantWomenWithAnemia -0.54779915  0.14731519  -3.7186 0.0002062 ***
## IncomePerCapita          0.00048670  0.00018528   2.6269 0.0086865 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    8589400
## Residual Sum of Squares: 177890
## R-Squared:      0.97929
## Adj. R-Squared: 0.97731
## F-statistic: 17911.3 on 5 and 1894 DF, p-value: < 2.22e-16
```

The inclusion of the lagged variable of the dependent variable is changing the statistical significance and the signs of some variables. Health expenditure and safe water access are statistically insignificant, and income per capita does not reflect the hypothesis stated at the beginning.

Third Table of Comparison

A table to compare the Pooled OLS model, the FE Two-way model, the RE Two-way model and the LDV model is constructed.

```
screenreg(list(   pooled_OLS,
                  twoway_effects,
                  RE_twoway,
                  ldv_model),

     custom.model.names = c("Pooled OLS",
                            "Twoway FE",
                            "Twoway RE",
                            "LDV Model"))
```

```
##
## ===========================================================================
##                           Pooled OLS   Twoway FE    Twoway RE    LDV Model
## ---------------------------------------------------------------------------
## (Intercept)                595.34 ***                757.05 ***
##                            (50.49)                    (1.05)
## SafeWaterAccess            -11.34 ***    -7.98 ***    -8.96 ***    -0.19
##                             (0.38)       (0.62)       (0.01)       (0.10)
## HealthExpenditure           29.46 ***    -6.83 ***    -5.24 ***    -0.38
##                             (2.18)       (1.61)       (0.02)       (0.25)
## PregnantWomenWithAnemia     11.64 ***     5.51 ***     7.01 ***    -0.55 ***
##                             (0.55)       (0.92)       (0.01)       (0.15)
## IncomePerCapita             -0.00         0.00 ***     0.00 ***     0.00 **
##                             (0.00)       (0.00)       (0.00)       (0.00)
## lag(MaternalMortality)                                               0.98 ***
##                                                                    (0.00)
## ---------------------------------------------------------------------------
## R^2                          0.69         0.16         0.65         0.98
## Adj. R^2                     0.69         0.09         0.64         0.98
## Num. obs.                 2076         2076         2076         2076
## s_idios                                                61.53
## s_id                                                  183.34
## s_time                                                  4.69
## ===========================================================================
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

The consolidated table compares the results of different models. The adjusted R-squared of the LDV model is much higher and therefore more significant, thanks to lagged dependent variable.

The following steps will be to determine the best model to be used.

The Poolability Test

Three tests are run comparing the three FE models with the Pooled OLS model. The null hypothesis is that the Pooled OLS is better than FE. In such case, the Pooled OLS would be selected as the best analysis model as there are no significant individual FE.

```r
pFtest(fixed_effects, pooled_OLS)
```

```
##
##  F test for individual effects
##
## data:  MaternalMortality ~ SafeWaterAccess + HealthExpenditure + PregnantWomenWithAnemia +  ...
## F = 117.11, df1 = 160, df2 = 1911, p-value < 2.2e-16
## alternative hypothesis: significant effects
```

The p-value is 0. The null hypothesis is rejected as there are significant individual FE. Therefore, the FE model is better than the Pooled OLS model.

```r
pFtest(twoway_effects, pooled_OLS)
```

```
##
##  F test for twoways effects
##
## data:  MaternalMortality ~ SafeWaterAccess + HealthExpenditure + PregnantWomenWithAnemia +  ...
## F = 106.11, df1 = 176, df2 = 1895, p-value < 2.2e-16
## alternative hypothesis: significant effects
```

Accounting for individual fixed effects and time effects. FE is better again.

p-values are low so we reject the null hypothesis, and we use the FE model instead of Pooled OLS. The panel effects are statistically significant because there is heterogeneity across countries.

The following step will be to check whether there are indeed any country fixed effects to begin with. According to the null hypothesis, there are no significant fixed effects.

```r
plmtest(fixed_effects, effect="individual")
```

```
##
##  Lagrange Multiplier Test - (Honda) for unbalanced panels
##
## data:  MaternalMortality ~ SafeWaterAccess + HealthExpenditure + PregnantWomenWithAnemia +  ...
## normal = 107.12, p-value < 2.2e-16
## alternative hypothesis: significant effects
```

The p-value of 0 suggests that the null hypothesis can be rejected and that there are indeed country fixed effects present in the Individual FE model.

The following test will check if there are time fixed effects in the Time FE model.

```r
plmtest(time_effects, effect="time")
```

```
##
##  Lagrange Multiplier Test - time effects (Honda) for unbalanced panels
##
## data:  MaternalMortality ~ SafeWaterAccess + HealthExpenditure + PregnantWomenWithAnemia +  ...
## normal = -2.1126, p-value = 0.9827
## alternative hypothesis: significant effects
```

The p-value is very high (0.9827) and thus we cannot reject the null hypothesis of no time FE. Hence, the selected model should not control for the time effects only.

Therefore, it is not suggested to control for time effects only, but what about controlling for both of them at the same time?

```
plmtest(twoway_effects, effect="twoways")
```

```
##
##  Lagrange Multiplier Test - two-ways effects (Honda) for unbalanced
##  panels
##
## data:  MaternalMortality ~ SafeWaterAccess + HealthExpenditure + PregnantWomenWithAnemia +  ...
## normal = 74.253, p-value < 2.2e-16
## alternative hypothesis: significant effects
```

The low p-value suggests to control for both individual and time effects in the "within" FE model.

The individual FE model result suggests to control for individual effects whereas according to the time effects model results, time effects alone are not significant. It is necessary to control for both of them at the same time (Two-way) or to simply use the individual FE model.

The Hausman Test

The Hausman test compares the FE and RE models, checking for the correlation between the regressors and the (individual or time) unobserved effects. The test does not give a definite answer to the question of which model to run, but it gives suggestions on the statistical significance of the RE but cannot be taken for granted because in the RE model, most of the coefficients are highly significant and therefore the Hausman test favours it.

```
phtest(twoway_effects, RE_twoway)
```

```
##
##  Hausman Test
##
## data:  MaternalMortality ~ SafeWaterAccess + HealthExpenditure + PregnantWomenWithAnemia +  ...
## chisq = 9.1067, df = 4, p-value = 0.05849
## alternative hypothesis: one model is inconsistent
```

The Hausman test suggests that the preferred model is the RE rather than the Two-way FE model as the null hypothesis cannot be rejected at the 5% significance level.

```
phtest(ldv_model, RE_twoway)
```

```
##
##  Hausman Test
##
## data:  MaternalMortality ~ lag(MaternalMortality) + SafeWaterAccess +  ...
## chisq = 14571, df = 4, p-value < 2.2e-16
## alternative hypothesis: one model is inconsistent
```

The Hausman test suggests that the preferred model is the LDV model which is more efficient at the 1% significance level.

The p-value is not low so we cannot reject the null hypothesis and conclude that RE is better than FE, but if the lagged dependent variable (LDV) is included, then the FE model is better than the RE model.

Serial Correlation Tests

As panel data has a time series dimension, we need to address the potential for serial correlation (AR) in the error term. Serial correlation in panel data will be tested with the Breusch-Godfrey test.

```
pbgtest(twoway_effects)
```

```
## 
##  Breusch-Godfrey/Wooldridge test for serial correlation in panel models
## 
## data:  MaternalMortality ~ SafeWaterAccess + HealthExpenditure + PregnantWomenWithAnemia +  ...
## chisq = 1493, df = 1, p-value < 2.2e-16
## alternative hypothesis: serial correlation in idiosyncratic errors
```

The null hypothesis of no serial correlation is rejected, there is evidence of autocorrelation in the residuals in the FE Two-way model.

```
pbgtest(RE_twoway)
```

```
## 
##  Breusch-Godfrey/Wooldridge test for serial correlation in panel models
## 
## data:  MaternalMortality ~ SafeWaterAccess + HealthExpenditure + PregnantWomenWithAnemia +  ...
## chisq = 1705.8, df = 1, p-value < 2.2e-16
## alternative hypothesis: serial correlation in idiosyncratic errors
```

The null hypothesis of no serial correlation is rejected, there is evidence of autocorrelation in the residuals in the RE Two-way model.

```
pbgtest(ldv_model)
```

```
## 
##  Breusch-Godfrey/Wooldridge test for serial correlation in panel models
## 
## data:  MaternalMortality ~ lag(MaternalMortality) + SafeWaterAccess +  ...
## chisq = 804.95, df = 1, p-value < 2.2e-16
## alternative hypothesis: serial correlation in idiosyncratic errors
```

The null hypothesis of no serial correlation is rejected, there is evidence of autocorrelation in the residuals in the LDV model.

There is evidence of serial correlation in the residuals in any test.

Corrections for Heteroskedasticity and Autocorrelation (HAC)

```
twoway_effects_hac = coeftest(twoway_effects,
                                               vcov = vcovHC(twoway_effects,
                                             method = "arellano",
                                             type   = "HC3"))

twoway_effects_hac
```

```
##
## t test of coefficients:
##
##                          Estimate Std. Error t value  Pr(>|t|)
## SafeWaterAccess        -7.9821605  2.0632821 -3.8687 0.0001131 ***
## HealthExpenditure      -6.8293578  7.6501470 -0.8927 0.3721262
## PregnantWomenWithAnemia  5.5100821  3.1196661  1.7662 0.0775164 .
## IncomePerCapita         0.0042492  0.0020284  2.0948 0.0363200 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The FE regression with two-way effects presents standard errors corrected for HC and AR. The standard errors are larger and some variables are not statistically significant because the standard error is larger causing the denominator of the t value to rise and the t value to decrease in absolute value causing the p-value to increase and becomes statistically insignificant.

Table of Comparison betweem Standard and Robust Errors

```
 screenreg(list(
               twoway_effects,
               twoway_effects_hac),

               custom.model.names = c("Twoway Fixed Effects",
                                      "Twoway Fixed Effects (HAC)"))
```

```
##
## ========================================================================
##                          Twoway Fixed Effects  Twoway Fixed Effects (HAC)
## ------------------------------------------------------------------------
## SafeWaterAccess             -7.98 ***            -7.98 ***
##                             (0.62)               (2.06)
## HealthExpenditure           -6.83 ***            -6.83
##                             (1.61)               (7.65)
## PregnantWomenWithAnemia      5.51 ***             5.51
##                             (0.92)               (3.12)
## IncomePerCapita              0.00 ***             0.00 *
##                             (0.00)               (0.00)
## ------------------------------------------------------------------------
## R^2                          0.16
## Adj. R^2                     0.09
```

```
## Num. obs.                 2076
## ==========================================================================
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

Cross Sectional Dependence (XSD) Test

In panel data with many countries, it is possible that there is a factor that is affecting all the countries at the same time but there is not any variable that controls for that factor. If the factor is controlled there is not any XSD but if it is not there is XSD. The following tests check for XSD in the Two-way FE, Two-way RE, and LDV models.

```
pcdtest(twoway_effects)
```

```
##
##  Pesaran CD test for cross-sectional dependence in panels
##
## data:  MaternalMortality ~ SafeWaterAccess + HealthExpenditure + PregnantWomenWithAnemia +     Income
## z = 30.574, p-value < 2.2e-16
## alternative hypothesis: cross-sectional dependence
```

The null hypothesis of no XSD is rejected, there is evidence of XSD.

```
pcdtest(RE_twoway)
```

```
##
##  Pesaran CD test for cross-sectional dependence in panels
##
## data:  MaternalMortality ~ SafeWaterAccess + HealthExpenditure + PregnantWomenWithAnemia +     Income
## z = 45.285, p-value < 2.2e-16
## alternative hypothesis: cross-sectional dependence
```

The null hypothesis of no XSD is rejected, there is evidence of XSD.

```
pcdtest(ldv_model)
```

```
##
##  Pesaran CD test for cross-sectional dependence in panels
##
## data:  MaternalMortality ~ lag(MaternalMortality) + SafeWaterAccess +     HealthExpenditure + Pregnan
## z = 47.974, p-value < 2.2e-16
## alternative hypothesis: cross-sectional dependence
```

The null hypothesis of no XSD is rejected, there is evidence of XSD.

Panel Corrected Standard Errors (PCSE)

If there is evidence of XSD, it is necessary to correct it through the Beck and Katz (1995) method.

```r
twoway_effects_pcse = coeftest(twoway_effects,
                                                    vcov = vcovBK(twoway_effects,
                                                    type = "HC3",
                                                 cluster = "group"))

twoway_effects_pcse
```

```
##
## t test of coefficients:
##
##                            Estimate Std. Error t value  Pr(>|t|)
## SafeWaterAccess          -7.9821605  2.2462144 -3.5536 0.0003893 ***
## HealthExpenditure        -6.8293578  4.5809657 -1.4908 0.1361773
## PregnantWomenWithAnemia  5.5100821  3.1068766  1.7735 0.0763045 .
## IncomePerCapita           0.0042492  0.0033367  1.2734 0.2030157
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Cross-sectional and Serial Correlation (SCC)

The results of the PCSE are sensitive to the ratio between the number of time periods in the dataset (T) and the total number of observations (N). The Cross-sectional and Serial Correlation (SCC) method by Driscoll and Kraay addresses the limitations of Beck and Katz's PCSE method and it is therefore preferred for obtaining Heteroskedasticity and Autocorrelation Consistent (HAC) errors that are also robust to cross-sectional dependence.

```r
twoway_effects_scc = coeftest(twoway_effects,
                                                   vcov = vcovSCC(twoway_effects,
                                                     type = "HC3",
                                                  cluster = "group"))

twoway_effects_scc
```

```
##
## t test of coefficients:
##
##                            Estimate Std. Error t value  Pr(>|t|)
## SafeWaterAccess          -7.9821605  1.9527163 -4.0877 4.538e-05 ***
## HealthExpenditure        -6.8293578  7.7192746 -0.8847   0.37642
## PregnantWomenWithAnemia  5.5100821  3.0548490  1.8037   0.07143 .
## IncomePerCapita           0.0042492  0.0019654  2.1620   0.03074 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The coefficient signs are as expected, apart from the income per capita, whose coefficient should be positive. Health expenditure is statistically insignificant because its effects might be incorporated into the other regressors.

Fourth Table of Comparison

The third table is complemented with the HAC, PCSE and SCC results.

```r
screenreg(list(

        pooled_OLS,
        RE_twoway,
        ldv_model,
        twoway_effects,
        twoway_effects_hac,
        twoway_effects_pcse,
        twoway_effects_scc),

 custom.model.names = c(

        "Pooled OLS",
        "Twoway RE",
        "LDV FE",
        "Twoway FE",
        "Arellano HAC FE",
        "Beck-Katz FE",
        "Driscoll-Kraay FE"))
```

```
##
## =====================================================================================================
##                           Pooled OLS    Twoway RE    LDV FE       Twoway FE    Arellano HAC FE  Beck-Ka
## -----------------------------------------------------------------------------------------------------
## (Intercept)                595.34 ***   757.05 ***
##                            (50.49)       (1.05)
## SafeWaterAccess            -11.34 ***    -8.96 ***    -0.19        -7.98 ***   -7.98 ***         -7.98 *
##                             (0.38)       (0.01)       (0.10)       (0.62)      (2.06)            (2.25)
## HealthExpenditure           29.46 ***    -5.24 ***    -0.38        -6.83 ***   -6.83             -6.83
##                             (2.18)       (0.02)       (0.25)       (1.61)      (7.65)            (4.58)
## PregnantWomenWithAnemia     11.64 ***     7.01 ***    -0.55 ***     5.51 ***    5.51              5.51
##                             (0.55)       (0.01)       (0.15)       (0.92)      (3.12)            (3.11)
## IncomePerCapita             -0.00         0.00 ***     0.00 **      0.00 ***    0.00 *            0.00
##                             (0.00)       (0.00)       (0.00)       (0.00)      (0.00)            (0.00)
## lag(MaternalMortality)                                 0.98 ***
##                                                       (0.00)
## -----------------------------------------------------------------------------------------------------
## R^2                           0.69         0.65         0.98         0.16
## Adj. R^2                      0.69         0.64         0.98         0.09
## Num. obs.                  2076         2076         2076         2076
## s_idios                                   61.53
## s_id                                     183.34
## s_time                                     4.69
## =====================================================================================================
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

Instrumental Variables, Endogeneity, and GMM Estimation

Baseline Non-IV Model

Fixed effects (FE) model with two-way effects (individual and time effects), with no instruments, is going to be our baseline model as it is the most efficient.

```
twoway_effects = plm(MaternalMortality ~ SafeWaterAccess + HealthExpenditure
                     + PregnantWomenWithAnemia + IncomePerCapita,

                        data = wdi,
                        index = c("Country", "Year"),
                        model = "within",
                        effect = "twoways")


summary(twoway_effects)
```

```
## Twoways effects Within Model
##
## Call:
## plm(formula = MaternalMortality ~ SafeWaterAccess + HealthExpenditure +
##     PregnantWomenWithAnemia + IncomePerCapita, data = wdi, effect = "twoways",
##     model = "within", index = c("Country", "Year"))
##
## Unbalanced Panel: n = 161, T = 1-17, N = 2076
##
## Residuals:
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -567.582  -14.088    0.000   12.432  539.814
##
## Coefficients:
##                            Estimate Std. Error  t-value  Pr(>|t|)
## SafeWaterAccess          -7.9821605  0.6187568 -12.9003 < 2.2e-16 ***
## HealthExpenditure        -6.8293578  1.6087862  -4.2450 2.291e-05 ***
## PregnantWomenWithAnemia   5.5100821  0.9245709   5.9596 3.008e-09 ***
## IncomePerCapita           0.0042492  0.0011730   3.6225 0.0002994 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    8589400
## Residual Sum of Squares: 7173800
## R-Squared:      0.16481
## Adj. R-Squared: 0.085483
## F-statistic: 93.4894 on 4 and 1895 DF, p-value: < 2.22e-16
```

The standard errors of the model are corrected for heteroscedasticity.

```
coeftest(twoway_effects, vcov = vcovHC, type = "HC1")
```

```
##
## t test of coefficients:
##
##                            Estimate Std. Error t value Pr(>|t|)
## SafeWaterAccess          -7.9821605  2.0522964 -3.8894 0.000104 ***
## HealthExpenditure        -6.8293578  7.6138500 -0.8970 0.369851
## PregnantWomenWithAnemia   5.5100821  3.1095656  1.7720 0.076559 .
## IncomePerCapita           0.0042492  0.0020134  2.1104 0.034952 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
coeftest(twoway_effects, vcov = vcovHC, type = "HC2")
```

```
##
## t test of coefficients:
##
##                          Estimate Std. Error t value  Pr(>|t|)
## SafeWaterAccess         -7.9821605  2.0567779 -3.8809 0.0001076 ***
## HealthExpenditure       -6.8293578  7.6282833 -0.8953 0.3707575
## PregnantWomenWithAnemia  5.5100821  3.1131044  1.7700 0.0768939 .
## IncomePerCapita          0.0042492  0.0020199  2.1037 0.0355396 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
coeftest(twoway_effects, vcov = vcovHC, type = "HC3")
```

```
##
## t test of coefficients:
##
##                          Estimate Std. Error t value  Pr(>|t|)
## SafeWaterAccess         -7.9821605  2.0632821 -3.8687 0.0001131 ***
## HealthExpenditure       -6.8293578  7.6501470 -0.8927 0.3721262
## PregnantWomenWithAnemia  5.5100821  3.1196661  1.7662 0.0775164 .
## IncomePerCapita          0.0042492  0.0020284  2.0948 0.0363200 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
coeftest(twoway_effects, vcov = vcovHC, type = "HC4")
```

```
##
## t test of coefficients:
##
##                          Estimate Std. Error t value  Pr(>|t|)
## SafeWaterAccess         -7.9821605  2.0726356 -3.8512 0.0001214 ***
## HealthExpenditure       -6.8293578  7.6754530 -0.8898 0.3737044
## PregnantWomenWithAnemia  5.5100821  3.1249081  1.7633 0.0780147 .
## IncomePerCapita          0.0042492  0.0020423  2.0806 0.0376037 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Two-stage Least Squares (2SLS) with External IV Model

Suppose there are reasons to believe that one of the regressors is not strictly exogenous. In this case, it would be necessary to find an instrumental variable (IV) for it. Endogeneity makes OLS estimates biased and inconsistent, and therefore must be remedied. The instrument must be correlated with the regressor but not with the error (it must be exogenous). Also, the instrument must affect the dependent variable only indirectly via the regressor, and not directly. It is necessary to have at least the same number of instruments as endogenous regressors, or maybe more IV than endogenous regressors, but not fewer instruments than endogenous regressors. The 2SLS estimator is consistent and normally distributed when the sample size is large. Suppose "IncomePerCapita" is endogenous to "MaternalMortality" (endogeneity problem). It is necessary to find an IV for "IncomePerCapita". To see how 2SLS works, an external IV that is in the dataset but not already in the model will be chosen. There is an endogenous variable on the right side of the equation which is causing an endogeneity problem. The instrument must be correlated with "IncomePerCapita" but uncorrelated with the error term, hence it must be an exogenous variable. "OilRents" can be used as an IV for "IncomePerCapita". In 2SLS regression, all of the instrumental variables are used to estimate all of the regression coefficients in the model.

```
Two_Stage_IV = plm(MaternalMortality ~ SafeWaterAccess +
                     HealthExpenditure + PregnantWomenWithAnemia
                   + IncomePerCapita

                                 | . - IncomePerCapita + OilRents,

                      data = wdi,
                     index = c("Country", "Year"),
                     model = "within",
                    effect = "twoways",
               inst.method = "bvk" )

summary(Two_Stage_IV)
```

```
## Twoways effects Within Model
## Instrumental variable estimation
##
## Call:
## plm(formula = MaternalMortality ~ SafeWaterAccess + HealthExpenditure +
##     PregnantWomenWithAnemia + IncomePerCapita | . - IncomePerCapita +
##     OilRents, data = wdi, effect = "twoways", model = "within",
##     inst.method = "bvk", index = c("Country", "Year"))
##
## Unbalanced Panel: n = 161, T = 1-17, N = 2068
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -547.7512  -26.0462   -2.0359   24.4174  555.7087
##
## Coefficients:
##                           Estimate Std. Error z-value  Pr(>|z|)
## SafeWaterAccess         -13.152020   2.261478 -5.8157 6.039e-09 ***
## HealthExpenditure        -5.773862   2.006924 -2.8770  0.004015 **
## PregnantWomenWithAnemia  10.195555   2.228458  4.5752 4.759e-06 ***
## IncomePerCapita          -0.029929   0.014194 -2.1085  0.034987 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    8546800
## Residual Sum of Squares: 10314000
## R-Squared:      0.032511
## Adj. R-Squared: -0.059778
## Chisq: 255.123 on 4 DF, p-value: < 2.22e-16
```

The following step will be to compute heteroskedasticity-robust standard errors.

```
coeftest(Two_Stage_IV, vcov = vcovHC, type = "HC1")
```

```
##
## t test of coefficients:
##
##                           Estimate Std. Error t value  Pr(>|t|)
## SafeWaterAccess         -13.152020   3.902475 -3.3702 0.0007663 ***
## HealthExpenditure        -5.773862   8.241739 -0.7006 0.4836618
## PregnantWomenWithAnemia  10.195555   4.737262  2.1522 0.0315078 *
## IncomePerCapita          -0.029929   0.017608 -1.6997 0.0893515 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
coeftest(Two_Stage_IV, vcov = vcovHC, type = "HC2")
```

```
##
## t test of coefficients:
##
##                           Estimate Std. Error t value  Pr(>|t|)
## SafeWaterAccess         -13.152020   3.909095 -3.3645 0.0007823 ***
## HealthExpenditure        -5.773862   8.256858 -0.6993 0.4844628
## PregnantWomenWithAnemia  10.195555   4.743896  2.1492 0.0317458 *
## IncomePerCapita          -0.029929   0.017654 -1.6953 0.0901834 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
coeftest(Two_Stage_IV, vcov = vcovHC, type = "HC3")
```

```
##
## t test of coefficients:
##
##                           Estimate Std. Error t value  Pr(>|t|)
## SafeWaterAccess         -13.152020   3.919606 -3.3554 0.0008081 ***
## HealthExpenditure        -5.773862   8.280055 -0.6973 0.4856874
## PregnantWomenWithAnemia  10.195555   4.755213  2.1441 0.0321539 *
## IncomePerCapita          -0.029929   0.017718 -1.6892 0.0913457 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
coeftest(Two_Stage_IV, vcov = vcovHC, type = "HC4")
```

```
##
## t test of coefficients:
##
##                           Estimate Std. Error t value  Pr(>|t|)
## SafeWaterAccess         -13.152020   3.932790 -3.3442 0.0008414 ***
## HealthExpenditure        -5.773862   8.306378 -0.6951 0.4870708
## PregnantWomenWithAnemia  10.195555   4.766783  2.1389 0.0325738 *
## IncomePerCapita          -0.029929   0.017816 -1.6799 0.0931411 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

comments

The GMM Model

The next step will be the estimation of the GMM model with lagged regressors as internal instruments. The generalized method of moments (GMM) is mainly used in panel data econometrics to estimate dynamic models with a lagged endogenous variable. In a GMM estimation, there are "normal instruments" and "GMM instruments". GMM instruments are indicated in the second part of the formula. The variables of the model that are not used as GMM instruments are used as normal instruments, with the same lag structure.

```
detach("package:dplyr", unload=TRUE)

                  GMM = pgmm(         MaternalMortality ~

                                      lag(MaternalMortality,       1:1) +
                                      lag(SafeWaterAccess,         0:2) +
                                      lag(HealthExpenditure,       0:2) +
                                      lag(PregnantWomenWithAnemia, 0:2) +
                                      lag(IncomePerCapita,         0:2)

                                    | lag(MaternalMortality,       2:6),

                            data = wdi,
                           index = c("Country", "Year"),
                           model = "onestep",
                          effect = "individual"    )

      summary(GMM)
```

```
## Oneway (individual) effect One-step model Difference GMM
##
## Call:
## pgmm(formula = MaternalMortality ~ lag(MaternalMortality, 1:1) +
##     lag(SafeWaterAccess, 0:2) + lag(HealthExpenditure, 0:2) +
##     lag(PregnantWomenWithAnemia, 0:2) + lag(IncomePerCapita,
##     0:2) | lag(MaternalMortality, 2:6), data = wdi, effect = "individual",
##     model = "onestep", index = c("Country", "Year"))
##
## Balanced Panel: n = 192, T = 56, N = 10752
##
## Number of Observations Used: 1650
## Residuals:
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -118.9926    0.0000    0.0000   -0.1755    0.0000  187.9692
##
## Coefficients:
##                                      Estimate  Std. Error z-value Pr(>|z|)
## lag(MaternalMortality, 1:1)         8.0389e-01  4.1741e-02 19.2590  < 2e-16 ***
## lag(SafeWaterAccess, 0:2)0         -1.7105e+02  8.7708e+01 -1.9502  0.05115 .
## lag(SafeWaterAccess, 0:2)1          5.6745e+01  4.6364e+01  1.2239  0.22099
## lag(SafeWaterAccess, 0:2)2          1.1206e+02  9.8676e+01  1.1357  0.25610
## lag(HealthExpenditure, 0:2)0       -6.2810e-01  7.6406e-01 -0.8221  0.41105
## lag(HealthExpenditure, 0:2)1       -2.2524e+00  1.3496e+00 -1.6689  0.09514 .
## lag(HealthExpenditure, 0:2)2       -1.5289e+00  1.6327e+00 -0.9364  0.34907
## lag(PregnantWomenWithAnemia, 0:2)0  3.1829e+01  3.0447e+01  1.0454  0.29585
```

```
## lag(PregnantWomenWithAnemia, 0:2)1  2.2681e+01  5.1436e+01  0.4410  0.65925
## lag(PregnantWomenWithAnemia, 0:2)2 -5.6499e+01  2.9085e+01 -1.9425  0.05207 .
## lag(IncomePerCapita, 0:2)0         -1.0940e-03  5.1526e-04 -2.1232  0.03374 *
## lag(IncomePerCapita, 0:2)1         -9.3662e-04  4.0254e-04 -2.3268  0.01998 *
## lag(IncomePerCapita, 0:2)2         -6.9003e-04  4.1353e-04 -1.6686  0.09519 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Sargan test: chisq(258) = 103.895 (p-value = 1)
## Autocorrelation test (1): normal = 0.4657612 (p-value = 0.64139)
## Autocorrelation test (2): normal = 0.2775533 (p-value = 0.78136)
## Wald test for coefficients: chisq(13) = 826.4293 (p-value = < 2.22e-16)
```

There are no problems with the instruments according to the Sargan test, as the p-value is greater than 0.25.

GMM with Logs and Lags of Logs

```
        GMM.logs = pgmm(        log(MaternalMortality) ~

                                lag(log(MaternalMortality),       1:1) +
                                lag(log(SafeWaterAccess),         0:2) +
                                lag(log(HealthExpenditure),       0:2) +
                                lag(log(PregnantWomenWithAnemia), 0:2) +
                                lag(log(IncomePerCapita),         0:2)

                              | lag(log(MaternalMortality),       2:6),

                     data = wdi,
                    index = c("Country", "Year"),
                    model = "onestep",
                   effect = "individual"    )

    summary(GMM.logs)
```

```
## Oneway (individual) effect One-step model Difference GMM
##
## Call:
## pgmm(formula = log(MaternalMortality) ~ lag(log(MaternalMortality),
##     1:1) + lag(log(SafeWaterAccess), 0:2) + lag(log(HealthExpenditure),
##     0:2) + lag(log(PregnantWomenWithAnemia), 0:2) + lag(log(IncomePerCapita),
##     0:2) | lag(log(MaternalMortality), 2:6), data = wdi, effect = "individual",
##     model = "onestep", index = c("Country", "Year"))
##
## Balanced Panel: n = 192, T = 56, N = 10752
##
## Number of Observations Used: 1641
## Residuals:
##       Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
## -0.5781694  0.0000000  0.0000000  0.0000416  0.0000000  0.6026289
##
## Coefficients:
```

```
##                                          Estimate Std. Error z-value Pr(>|z|)
## lag(log(MaternalMortality), 1:1)         0.9746478  0.0403901 24.1308  < 2e-16
## lag(log(SafeWaterAccess), 0:2)0         -0.7405508  1.2913218 -0.5735  0.56632
## lag(log(SafeWaterAccess), 0:2)1         -0.6020512  1.8824073 -0.3198  0.74910
## lag(log(SafeWaterAccess), 0:2)2          1.1995344  1.3596833  0.8822  0.37766
## lag(log(HealthExpenditure), 0:2)0        0.0226051  0.0142531  1.5860  0.11274
## lag(log(HealthExpenditure), 0:2)1       -0.0245323  0.0133555 -1.8369  0.06623
## lag(log(HealthExpenditure), 0:2)2        0.0228813  0.0134527  1.7009  0.08897
## lag(log(PregnantWomenWithAnemia), 0:2)0  0.0600413  0.6252816  0.0960  0.92350
## lag(log(PregnantWomenWithAnemia), 0:2)1 -0.2117659  1.2091179 -0.1751  0.86097
## lag(log(PregnantWomenWithAnemia), 0:2)2  0.1430010  0.6985612  0.2047  0.83780
## lag(log(IncomePerCapita), 0:2)0          0.0228533  0.0225493  1.0135  0.31083
## lag(log(IncomePerCapita), 0:2)1         -0.0310295  0.0240309 -1.2912  0.19662
## lag(log(IncomePerCapita), 0:2)2         -0.0019742  0.0198089 -0.0997  0.92061
##
## lag(log(MaternalMortality), 1:1)        ***
## lag(log(SafeWaterAccess), 0:2)0
## lag(log(SafeWaterAccess), 0:2)1
## lag(log(SafeWaterAccess), 0:2)2
## lag(log(HealthExpenditure), 0:2)0
## lag(log(HealthExpenditure), 0:2)1       .
## lag(log(HealthExpenditure), 0:2)2       .
## lag(log(PregnantWomenWithAnemia), 0:2)0
## lag(log(PregnantWomenWithAnemia), 0:2)1
## lag(log(PregnantWomenWithAnemia), 0:2)2
## lag(log(IncomePerCapita), 0:2)0
## lag(log(IncomePerCapita), 0:2)1
## lag(log(IncomePerCapita), 0:2)2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Sargan test: chisq(258) = 81.76918 (p-value = 1)
## Autocorrelation test (1): normal = -4.791843 (p-value = 1.6526e-06)
## Autocorrelation test (2): normal = 2.454765 (p-value = 0.014098)
## Wald test for coefficients: chisq(13) = 6295.532 (p-value = < 2.22e-16)
```

The regressors are statistically insignificant.

IV Models in a Consolidated Table

The models will be grouped side-by-side in a single consolidated table.

```
        screenreg(list(

                  twoway_effects,
                  Two_Stage_IV,
                  GMM,
                  GMM.logs
                               ),

         custom.model.names = c(

                  "FE two-way",
```

```
                    "Two_Stage_IV",
                    "GMM",
                    "GMM.logs"
                                 ))
```

```
##
## ==============================================================================================
##                                           FE two-way   Two_Stage_IV  GMM           GMM.logs
## ----------------------------------------------------------------------------------------------
## SafeWaterAccess                             -7.98 ***   -13.15 ***
##                                             (0.62)      (2.26)
## HealthExpenditure                           -6.83 ***    -5.77 **
##                                             (1.61)      (2.01)
## PregnantWomenWithAnemia                      5.51 ***    10.20 ***
##                                             (0.92)      (2.23)
## IncomePerCapita                              0.00 ***    -0.03 *
##                                             (0.00)      (0.01)
## lag(MaternalMortality, 1:1)                                            0.80 ***
##                                                                       (0.04)
## lag(SafeWaterAccess, 0:2)0                                          -171.05
##                                                                      (87.71)
## lag(SafeWaterAccess, 0:2)1                                            56.74
##                                                                      (46.36)
## lag(SafeWaterAccess, 0:2)2                                           112.06
##                                                                      (98.68)
## lag(HealthExpenditure, 0:2)0                                          -0.63
##                                                                       (0.76)
## lag(HealthExpenditure, 0:2)1                                          -2.25
##                                                                       (1.35)
## lag(HealthExpenditure, 0:2)2                                          -1.53
##                                                                       (1.63)
## lag(PregnantWomenWithAnemia, 0:2)0                                    31.83
##                                                                      (30.45)
## lag(PregnantWomenWithAnemia, 0:2)1                                    22.68
##                                                                      (51.44)
## lag(PregnantWomenWithAnemia, 0:2)2                                   -56.50
##                                                                      (29.09)
## lag(IncomePerCapita, 0:2)0                                            -0.00 *
##                                                                       (0.00)
## lag(IncomePerCapita, 0:2)1                                            -0.00 *
##                                                                       (0.00)
## lag(IncomePerCapita, 0:2)2                                            -0.00
##                                                                       (0.00)
## lag(log(MaternalMortality), 1:1)                                                     0.97 ***
##                                                                                     (0.04)
## lag(log(SafeWaterAccess), 0:2)0                                                     -0.74
##                                                                                     (1.29)
## lag(log(SafeWaterAccess), 0:2)1                                                     -0.60
##                                                                                     (1.88)
## lag(log(SafeWaterAccess), 0:2)2                                                      1.20
##                                                                                     (1.36)
## lag(log(HealthExpenditure), 0:2)0                                                    0.02
##                                                                                     (0.01)
```

```
## lag(log(HealthExpenditure), 0:2)1                                                               -0.02
##                                                                                                 (0.01)
## lag(log(HealthExpenditure), 0:2)2                                                                0.02
##                                                                                                 (0.01)
## lag(log(PregnantWomenWithAnemia), 0:2)0                                                          0.06
##                                                                                                 (0.63)
## lag(log(PregnantWomenWithAnemia), 0:2)1                                                         -0.21
##                                                                                                 (1.21)
## lag(log(PregnantWomenWithAnemia), 0:2)2                                                          0.14
##                                                                                                 (0.70)
## lag(log(IncomePerCapita), 0:2)0                                                                  0.02
##                                                                                                 (0.02)
## lag(log(IncomePerCapita), 0:2)1                                                                 -0.03
##                                                                                                 (0.02)
## lag(log(IncomePerCapita), 0:2)2                                                                 -0.00
##                                                                                                 (0.02)
## ---------------------------------------------------------------------------------------------------
## R^2                                         0.16         0.03
## Adj. R^2                                    0.09        -0.06
## Num. obs.                                2076         2068          10752         10752
## n                                                                     192           192
## T                                                                      56            56
## Num. obs. used                                                       1650          1641
## Sargan Test: chisq                                                    103.90         81.77
## Sargan Test: df                                                       258.00        258.00
## Sargan Test: p-value                                                    1.00          1.00
## Wald Test Coefficients: chisq                                         826.43       6295.53
## Wald Test Coefficients: df                                             13            13
## Wald Test Coefficients: p-value                                          0.00          0.00
## ===================================================================================================
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

Conclusion

The 2SLS is the best model as all variables are significant and have the correct sign. Maternal mortality decreases with more access to safe water, higher health expenditure, and higher income per capita. On the other hand, maternal mortality increases with an increase in pregnant women with anemia.