Component 2, Stage 2

Sevastian Sanchez

2025-09-20

```
knitr::opts_chunk$set(
  echo = TRUE, warning = FALSE, message = FALSE,
  tidy = TRUE, tidy.opts = list(width.cutoff = 60)
)
```

1 Set up

```
# set working directory
setwd("~/Documents/GitHub/QMSS_Thesis_Sanchez")

# load libraries/packages
source("packages.R")

# load data
source("Comp2_panel_wrangling.R")
# select path = 'data/Main CSV
# Outputs/merged_cleaned_sdg.csv'

# select relevant variables and arrange data
panel_data1 <- panel_data %>%
    dplyr::select(country_name, country_code, year, sdg_overall,
        spi_comp, di_score, log_gdppc, income_level_recoded) %>%
    dplyr::arrange(country_code, year)

# how many countries
length(unique(panel_data1$country_code))
```

```
## [1] 162
```

```
# check dimensions
dim(panel_data1)
```

```
## [1] 1296    8
```

1.1 Converting to panel data frame

```r
panel_data <- pdata.frame(panel_data1, index = c("country_code",
    "year"))
pdim(panel_data)  # check panel dimensions
```

```
## Balanced Panel: n = 162, T = 8, N = 1296
```

2 2.1) POLS SDG ~ SPI [Stage 2]

```r
# Contemporaneous Effect: SDG ~ SPI + DI
ols_sdg_spi <- plm(formula = sdg_overall ~ spi_comp + di_score + log_gdppc
                   + factor(income_level_recoded) + factor(year),
               model = "pooling",
               data = panel_data)
summary(ols_sdg_spi, vcov = vcovHC(ols_sdg_spi, cluster = "group", type = "HC1"))
```

```
## Pooling Model
##
## Note: Coefficient variance-covariance matrix supplied: vcovHC(ols_sdg_spi, cluster = "group", type =
##
## Call:
## plm(formula = sdg_overall ~ spi_comp + di_score + log_gdppc +
##     factor(income_level_recoded) + factor(year), data = panel_data,
##     model = "pooling")
##
## Unbalanced Panel: n = 155, T = 6-8, N = 1234
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -11.96220  -2.86494   0.11405   2.84443  13.98364
##
## Coefficients:
##                                 Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)                    27.473206   4.495280  6.1116 1.325e-09 ***
## spi_comp                        0.284667   0.033946  8.3859 < 2.2e-16 ***
## di_score                        0.291655   0.256248  1.1382 0.2552714
## log_gdppc                       1.600463   0.734638  2.1786 0.0295543 *
## factor(income_level_recoded)1   4.718372   1.236004  3.8174 0.0001416 ***
## factor(income_level_recoded)2   8.481679   1.740942  4.8719 1.251e-06 ***
## factor(income_level_recoded)3   7.916414   2.577748  3.0711 0.0021802 **
## factor(year)2017               -0.276635   0.153814 -1.7985 0.0723446 .
## factor(year)2018               -0.831429   0.251746 -3.3027 0.0009855 ***
## factor(year)2019               -0.531674   0.271031 -1.9617 0.0500276 .
## factor(year)2020               -0.710819   0.363210 -1.9570 0.0505695 .
## factor(year)2021               -1.964981   0.495922 -3.9623 7.854e-05 ***
## factor(year)2022               -1.806874   0.480822 -3.7579 0.0001795 ***
## factor(year)2023               -2.226746   0.492036 -4.5256 6.612e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Total Sum of Squares:    124070
## Residual Sum of Squares: 24004
## R-Squared:      0.80654
## Adj. R-Squared: 0.80448
## F-statistic: 77.2465 on 13 and 154 DF, p-value: < 2.22e-16
```

```r
# Adding Lag1: SPI ~ DI
ols_sdg_spi_L1 <- plm(formula = sdg_overall ~ spi_comp + plm::lag(spi_comp, 1)
                      + di_score + plm::lag(di_score, 1)
                      + log_gdppc
                      #+ factor(income_level_recoded)
                      + factor(year),
              model = "pooling",
              data = panel_data)
summary(ols_sdg_spi_L1, vcov = vcovHC(ols_sdg_spi_L1, cluster = "group", type = "HC1"))
```

```
## Pooling Model
##
## Note: Coefficient variance-covariance matrix supplied: vcovHC(ols_sdg_spi_L1, cluster = "group", typ
##
## Call:
## plm(formula = sdg_overall ~ spi_comp + plm::lag(spi_comp, 1) +
##     di_score + plm::lag(di_score, 1) + log_gdppc + factor(year),
##     data = panel_data, model = "pooling")
##
## Unbalanced Panel: n = 155, T = 5-7, N = 1079
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -12.088589  -3.084727  -0.060242   3.098490  13.397951
##
## Coefficients:
##                          Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)           19.291677   2.536404  7.6059 6.190e-14 ***
## spi_comp               0.078897   0.066955  1.1784 0.2389142
## plm::lag(spi_comp, 1)  0.211937   0.060640  3.4950 0.0004935 ***
## di_score               0.545687   0.675244  0.8081 0.4191941
## plm::lag(di_score, 1) -0.374345   0.690778 -0.5419 0.5879885
## log_gdppc              3.279639   0.472111  6.9468 6.489e-12 ***
## factor(year)2018      -0.514305   0.138793 -3.7056 0.0002218 ***
## factor(year)2019      -0.636962   0.237108 -2.6864 0.0073353 **
## factor(year)2020      -0.227394   0.281357 -0.8082 0.4191527
## factor(year)2021      -1.243651   0.446653 -2.7844 0.0054579 **
## factor(year)2022      -2.278122   0.468727 -4.8602 1.348e-06 ***
## factor(year)2023      -2.273578   0.465163 -4.8877 1.177e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    107440
## Residual Sum of Squares: 23790
## R-Squared:      0.77858
## Adj. R-Squared: 0.77629
## F-statistic: 70.2306 on 11 and 154 DF, p-value: < 2.22e-16
```

```r
# Adding Lag2: SPI ~ DI
ols_sdg_spi_L2 <- plm(formula = sdg_overall ~ spi_comp + plm::lag(spi_comp, 1)
                      + plm::lag(spi_comp, 2) + di_score + plm::lag(di_score, 1)
                      + plm::lag(di_score, 2) + log_gdppc
                      #+ factor(income_level_recoded)
                      + factor(year),
                model = "pooling",
                data = panel_data)
summary(ols_sdg_spi_L2, vcov = vcovHC(ols_sdg_spi_L2, cluster = "group", type = "HC1"))
```

```
## Pooling Model
##
## Note: Coefficient variance-covariance matrix supplied: vcovHC(ols_sdg_spi_L2, cluster = "group", typ
##
## Call:
## plm(formula = sdg_overall ~ spi_comp + plm::lag(spi_comp, 1) +
##     plm::lag(spi_comp, 2) + di_score + plm::lag(di_score, 1) +
##     plm::lag(di_score, 2) + log_gdppc + factor(year), data = panel_data,
##     model = "pooling")
##
## Unbalanced Panel: n = 155, T = 4-6, N = 924
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -11.18756  -3.03893  -0.19835   3.09998  12.60166
##
## Coefficients:
##                          Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)            19.224554   2.553358  7.5291 1.228e-13 ***
## spi_comp                0.117994   0.075368  1.5656 0.1177960
## plm::lag(spi_comp, 1)   0.013656   0.031069  0.4395 0.6603797
## plm::lag(spi_comp, 2)   0.165954   0.060999  2.7206 0.0066406 **
## di_score                0.475898   0.619350  0.7684 0.4424588
## plm::lag(di_score, 1)   0.064283   0.353793  0.1817 0.8558606
## plm::lag(di_score, 2)  -0.404698   0.698480 -0.5794 0.5624640
## log_gdppc               3.235028   0.472059  6.8530 1.331e-11 ***
## factor(year)2019       -0.031128   0.190694 -0.1632 0.8703706
## factor(year)2020       -0.057259   0.252872 -0.2264 0.8209142
## factor(year)2021       -0.830884   0.374317 -2.2197 0.0266823 *
## factor(year)2022       -1.257564   0.373981 -3.3626 0.0008042 ***
## factor(year)2023       -2.251831   0.462339 -4.8705 1.312e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    91386
## Residual Sum of Squares: 19961
## R-Squared:      0.78158
## Adj. R-Squared: 0.7787
## F-statistic: 62.5704 on 12 and 154 DF, p-value: < 2.22e-16
```

2.1 POLS Scatterplots

General relationship between SDG and Main Xs: SPI & DI

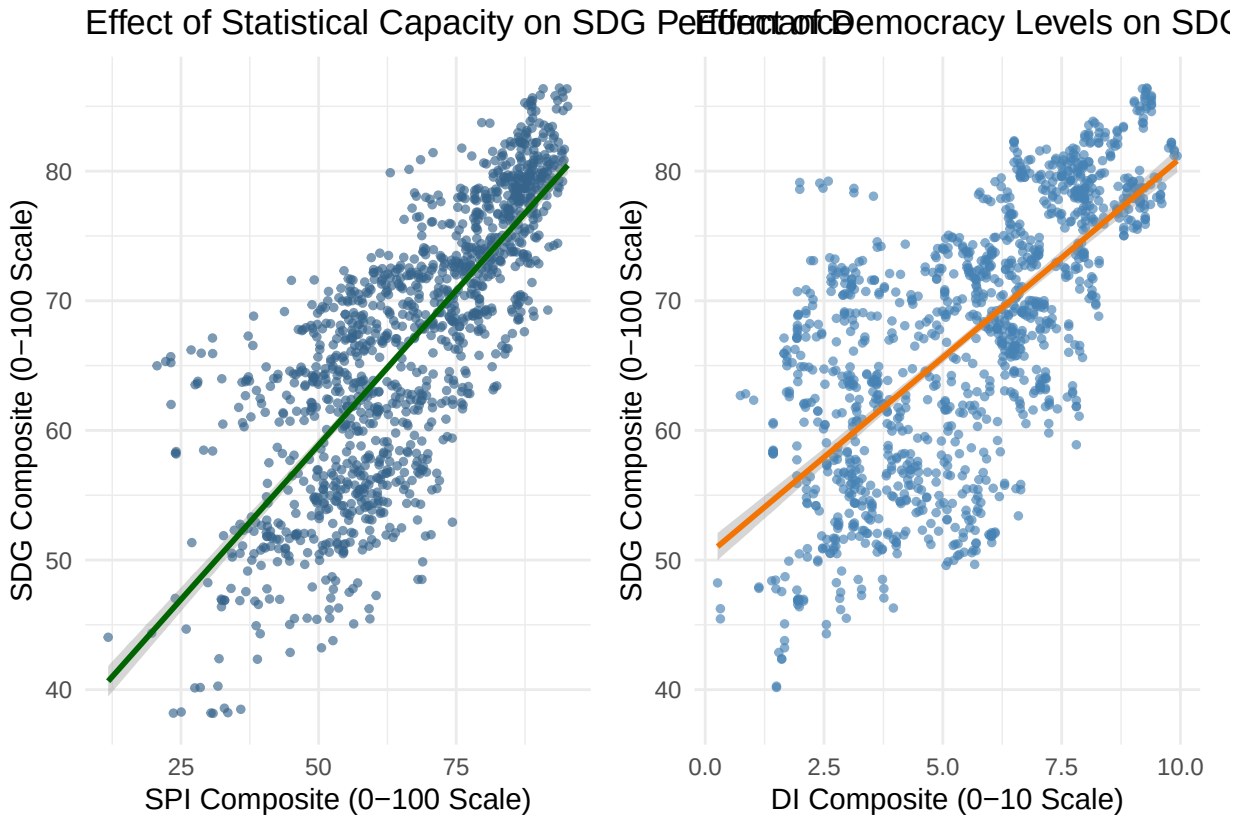
```r
# Contemporaneous Relationship: SDG ~ SPI
sdg_spi_s2_scatter <- ggplot(panel_data1, aes(x = spi_comp, y = sdg_overall)) +
    geom_point(color = "steelblue4", size = 1, alpha = 0.65) +
    geom_smooth(method = "lm", se = TRUE, color = "darkgreen",
        size = 1) + labs(title = "Effect of Statistical Capacity on SDG Performance",
    x = "SPI Composite (0-100 Scale)", y = "SDG Composite (0-100 Scale)") +
    theme(plot.title = element_text(size = 14)) + theme_minimal()

# Save to specific folder
# ggsave('component_2/figures/stage2/sdg_spi_s2_scatterplot.png',
# sdg_spi_s2_scatter, width = 8, height = 6)

# Contemporaneous Relationship: SDG ~ DI
sdg_di_s2_scatter <- ggplot(panel_data1, aes(x = di_score, y = sdg_overall)) +
    geom_point(color = "steelblue", size = 1, alpha = 0.65) +
    geom_smooth(method = "lm", se = TRUE, color = "#f27405",
        size = 1) + labs(title = "Effect of Democracy Levels on SDG Performance",
    x = "DI Composite (0-10 Scale)", y = "SDG Composite (0-100 Scale)") +
    theme(plot.title = element_text(size = 14)) + theme_minimal()

# Save to specific folder
# ggsave('component_2/figures/stage2/sdg_di_s2_scatterplot.png',
# sdg_di_s2_scatter, width = 8, height = 6)


# side by side comparison using patchwork
library(patchwork)
sdg_spi_s2_scatter + sdg_di_s2_scatter + plot_layout(ncol = 2)
```




```
# ggsave('component_2/figures/stage2/s2_scatterplots.png',
# width = 12, height = 8)
```

2.2 Stargazer Table for POLS Models

3 2.2) First Difference [Stage 2]

```
# Contemporaneous Effect: SDG ~ SPI + DI
fd_sdg_spi <- plm(formula = sdg_overall ~ spi_comp + di_score
                  + log_gdppc,
                  #+ factor(income_level_recoded),
                model = "fd",
                data = fd_data)
summary(fd_sdg_spi, vcov = vcovHC(fd_sdg_spi, cluster = "group", type = "HC1"))
```

```
## Oneway (individual) effect First-Difference Model
##
## Note: Coefficient variance-covariance matrix supplied: vcovHC(fd_sdg_spi, cluster = "group", type =
##
## Call:
## plm(formula = sdg_overall ~ spi_comp + di_score + log_gdppc,
##     data = fd_data, model = "fd")
##
## Unbalanced Panel: n = 155, T = 6-8, N = 1234
```

```
## Observations used in estimation: 1079
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -1.936438 -0.317256 -0.048017  0.246820  3.136182
##
## Coefficients:
##               Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)  0.3059277  0.0194780 15.7063 < 2.2e-16 ***
## spi_comp     0.0158925  0.0059871  2.6545  0.008061 **
## di_score     0.0364382  0.0837756  0.4350  0.663686
## log_gdppc   -0.2235660  0.2184315 -1.0235  0.306299
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    348.2
## Residual Sum of Squares: 344.9
## R-Squared:      0.009467
## Adj. R-Squared: 0.0067027
## F-statistic: 2.49471 on 3 and 154 DF, p-value: 0.062049
```

```
# Adding Lag1: SPI ~ DI
fd_sdg_spi_L1 <- plm(formula = sdg_overall ~ spi_comp + plm::lag(spi_comp, 1)
                     + di_score + plm::lag(di_score, 1)
                     + log_gdppc,
                     #+ factor(income_level_recoded),
                model = "fd",
                data = fd_data)
summary(fd_sdg_spi_L1, vcov = vcovHC(fd_sdg_spi_L1, cluster = "group", type = "HC1"))
```

```
## Oneway (individual) effect First-Difference Model
##
## Note: Coefficient variance-covariance matrix supplied: vcovHC(fd_sdg_spi_L1, cluster = "group", type
##
## Call:
## plm(formula = sdg_overall ~ spi_comp + plm::lag(spi_comp, 1) +
##     di_score + plm::lag(di_score, 1) + log_gdppc, data = fd_data,
##     model = "fd")
##
## Unbalanced Panel: n = 155, T = 5-7, N = 1079
## Observations used in estimation: 924
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -1.932564 -0.310496 -0.047395  0.249050  3.106300
##
## Coefficients:
##                          Estimate Std. Error t-value Pr(>|t|)
## (Intercept)             0.2869939  0.0248546 11.5469  < 2e-16 ***
## spi_comp                0.0111140  0.0064228  1.7304  0.08390 .
## plm::lag(spi_comp, 1)   0.0075453  0.0060299  1.2513  0.21114
## di_score                0.0655585  0.0838080  0.7822  0.43427
## plm::lag(di_score, 1)   0.1369567  0.0735237  1.8628  0.06282 .
## log_gdppc              -0.3168694  0.2451118 -1.2928  0.19642
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    303.64
## Residual Sum of Squares: 300.06
## R-Squared:      0.011763
## Adj. R-Squared: 0.0063802
## F-statistic: 1.60127 on 5 and 154 DF, p-value: 0.16294
```

```
# Adding Lag2: SPI ~ DI
fd_sdg_spi_L2 <- plm(formula = sdg_overall ~ spi_comp
                     + plm::lag(spi_comp, 1) + plm::lag(spi_comp, 2)
                     + di_score + plm::lag(di_score, 1) + plm::lag(di_score, 2)
                     + log_gdppc,
                     #+ factor(income_level_recoded),
                model = "fd",
                data = fd_data)
summary(fd_sdg_spi_L2, vcov = vcovHC(fd_sdg_spi_L2, cluster = "group", type = "HC1"))
```

```
## Oneway (individual) effect First-Difference Model
##
## Note: Coefficient variance-covariance matrix supplied: vcovHC(fd_sdg_spi_L2, cluster = "group", type
##
## Call:
## plm(formula = sdg_overall ~ spi_comp + plm::lag(spi_comp, 1) +
##     plm::lag(spi_comp, 2) + di_score + plm::lag(di_score, 1) +
##     plm::lag(di_score, 2) + log_gdppc, data = fd_data, model = "fd")
##
## Unbalanced Panel: n = 155, T = 4-6, N = 924
## Observations used in estimation: 769
##
## Residuals:
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -1.85910 -0.30627 -0.03719  0.25491  2.90944
##
## Coefficients:
##                          Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)             0.2363339  0.0283602  8.3333 3.658e-16 ***
## spi_comp                0.0170180  0.0070463  2.4152   0.01596 *
## plm::lag(spi_comp, 1)   0.0063430  0.0057420  1.1047   0.26965
## plm::lag(spi_comp, 2)   0.0132665  0.0059627  2.2249   0.02638 *
## di_score                0.0494996  0.0961530  0.5148   0.60684
## plm::lag(di_score, 1)   0.1845462  0.0756324  2.4400   0.01491 *
## plm::lag(di_score, 2) -0.1754561  0.1254654 -1.3984   0.16239
## log_gdppc             -0.4578589  0.2383173 -1.9212   0.05508 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    229.74
## Residual Sum of Squares: 222.38
## R-Squared:      0.032023
## Adj. R-Squared: 0.023119
## F-statistic: 2.16398 on 7 and 154 DF, p-value: 0.040328
```

3.1 Stargazer Table for FD Models

4 2.3) Fixed Effects [Stage 2]

```
# Contemporaneous Effect: SDG ~ SPI + DI
fe_sdg_spi <- plm(formula = sdg_overall ~ spi_comp + di_score
                  + log_gdppc
                  #+ factor(income_level_recoded)
                  + factor(year),
                  data = panel_data,
                  model = "within" #FE
)
summary(fe_sdg_spi, vcov = vcovHC(fe_sdg_spi, cluster = "group", type = "HC1")) # Robust SEs
```

```
## Oneway (individual) effect Within Model
##
## Note: Coefficient variance-covariance matrix supplied: vcovHC(fe_sdg_spi, cluster = "group", type =
##
## Call:
## plm(formula = sdg_overall ~ spi_comp + di_score + log_gdppc +
##     factor(year), data = panel_data, model = "within")
##
## Unbalanced Panel: n = 155, T = 6-8, N = 1234
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -2.8383110 -0.3014919 -0.0055603  0.3235414  3.4415657
##
## Coefficients:
##                   Estimate Std. Error t-value  Pr(>|t|)
## spi_comp          0.043690   0.014911  2.9300  0.003462 **
## di_score          0.096001   0.089973  1.0670  0.286215
## log_gdppc         0.412324   0.332591  1.2397  0.215346
## factor(year)2017 0.333735   0.054558  6.1170 1.335e-09 ***
## factor(year)2018 0.554422   0.093703  5.9168 4.415e-09 ***
## factor(year)2019 0.943784   0.099870  9.4502 < 2.2e-16 ***
## factor(year)2020 1.270535   0.108616 11.6975 < 2.2e-16 ***
## factor(year)2021 1.241928   0.167769  7.4026 2.700e-13 ***
## factor(year)2022 1.438483   0.171331  8.3959 < 2.2e-16 ***
## factor(year)2023 1.611548   0.188216  8.5622 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    1162.9
## Residual Sum of Squares: 447.51
## R-Squared:      0.61517
## Adj. R-Squared: 0.55614
## F-statistic: 57.0598 on 10 and 154 DF, p-value: < 2.22e-16
```

```
# Adding Lag1: SPI ~ DI
fe_sdg_spi_L1 <- plm(formula = sdg_overall ~ spi_comp + plm::lag(spi_comp, 1)
```

```
                    + di_score + plm::lag(di_score, 1)
                    + log_gdppc
                    #+ factor(income_level_recoded)
                    + factor(year),
                model = "within",
                data = panel_data)
summary(fe_sdg_spi_L1, vcov = vcovHC(fe_sdg_spi_L1, cluster = "group", type = "HC1"))
```

```
## Oneway (individual) effect Within Model
##
## Note: Coefficient variance-covariance matrix supplied: vcovHC(fe_sdg_spi_L1, cluster = "group", type
##
## Call:
## plm(formula = sdg_overall ~ spi_comp + plm::lag(spi_comp, 1) +
##     di_score + plm::lag(di_score, 1) + log_gdppc + factor(year),
##     data = panel_data, model = "within")
##
## Unbalanced Panel: n = 155, T = 5-7, N = 1079
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -2.658102 -0.266671 -0.014237  0.259776  3.149806
##
## Coefficients:
##                          Estimate Std. Error t-value  Pr(>|t|)
## spi_comp                0.0242798  0.0113188  2.1451 0.0322088 *
## plm::lag(spi_comp, 1)   0.0221931  0.0098094  2.2624 0.0239044 *
## di_score                0.2089459  0.0876375  2.3842 0.0173194 *
## plm::lag(di_score, 1) -0.0345650  0.1078151 -0.3206 0.7485902
## log_gdppc               0.3781954  0.3864323  0.9787 0.3279950
## factor(year)2018        0.2213588  0.0656081  3.3740 0.0007722 ***
## factor(year)2019        0.5649481  0.0797209  7.0866 2.748e-12 ***
## factor(year)2020        0.9354472  0.0896732 10.4317 < 2.2e-16 ***
## factor(year)2021        0.9611212  0.1371212  7.0093 4.654e-12 ***
## factor(year)2022        1.0380810  0.1651579  6.2854 5.060e-10 ***
## factor(year)2023        1.2559573  0.1724737  7.2820 7.098e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    740.6
## Residual Sum of Squares: 317.02
## R-Squared:      0.57195
## Adj. R-Squared: 0.49459
## F-statistic: 41.4617 on 11 and 154 DF, p-value: < 2.22e-16
```

```
# Adding Lag2: SPI ~ DI
fe_sdg_spi_L2 <- plm(formula = sdg_overall ~ spi_comp + plm::lag(spi_comp, 1)
                    + plm::lag(spi_comp, 2) + di_score + plm::lag(di_score, 1)
                    + plm::lag(di_score, 2) + log_gdppc
                    #+ factor(income_level_recoded)
                    + factor(year),
                model = "within",
                data = panel_data)
```

```r
summary(fe_sdg_spi_L2, vcov = vcovHC(fe_sdg_spi_L2, cluster = "group", type = "HC1"))
```

```
## Oneway (individual) effect Within Model
##
## Note: Coefficient variance-covariance matrix supplied: vcovHC(fe_sdg_spi_L2, cluster = "group", type
##
## Call:
## plm(formula = sdg_overall ~ spi_comp + plm::lag(spi_comp, 1) +
##     plm::lag(spi_comp, 2) + di_score + plm::lag(di_score, 1) +
##     plm::lag(di_score, 2) + log_gdppc + factor(year), data = panel_data,
##     model = "within")
##
## Unbalanced Panel: n = 155, T = 4-6, N = 924
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -2.4066416 -0.2350016 -0.0064308  0.2388494  2.8673678
##
## Coefficients:
##                          Estimate Std. Error t-value  Pr(>|t|)
## spi_comp                0.0343922  0.0120002  2.8660  0.004273 **
## plm::lag(spi_comp, 1)   0.0042770  0.0077243  0.5537  0.579945
## plm::lag(spi_comp, 2)   0.0155108  0.0088438  1.7539  0.079858 .
## di_score                0.1262507  0.0799246  1.5796  0.114611
## plm::lag(di_score, 1)   0.1752362  0.0968028  1.8102  0.070655 .
## plm::lag(di_score, 2)  -0.2359582  0.1278788 -1.8452  0.065403 .
## log_gdppc              -0.0010185  0.3868273 -0.0026  0.997900
## factor(year)2019        0.3409033  0.0450779  7.5625 1.147e-13 ***
## factor(year)2020        0.6307859  0.0778682  8.1007 2.186e-15 ***
## factor(year)2021        0.6912256  0.1080382  6.3980 2.757e-10 ***
## factor(year)2022        0.8508653  0.1324075  6.4261 2.314e-10 ***
## factor(year)2023        0.9597129  0.1660486  5.7797 1.094e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    440.52
## Residual Sum of Squares: 201.98
## R-Squared:      0.54148
## Adj. R-Squared: 0.44093
## F-statistic: 34.4067 on 12 and 154 DF, p-value: < 2.22e-16
```

5 Stargazer Table for FE Models

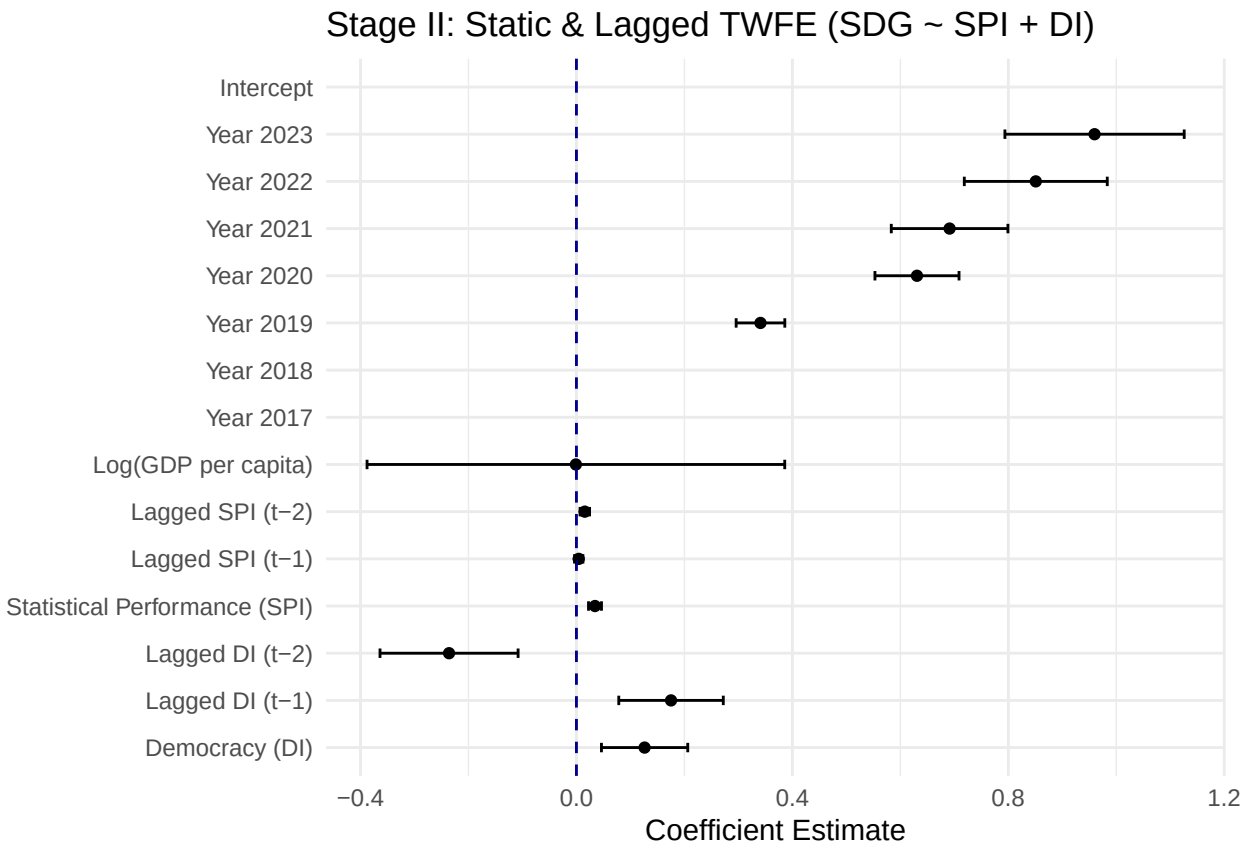
5.1 make a stargazer table of all Lag2 models for POLS, FD and FE

5.2 Fixed Effects Error Bar Visualization

```r
# Extract coefficients and robust standard errors from the
# FE model
coefs <- summary(fe_sdg_spi_L2, vcov = vcovHC(fe_sdg_spi_L2,
```

```r
    cluster = "group", type = "HC1"))$coefficients
# Create a data frame for visualization
coef_df <- data.frame(term = rownames(coefs), estimate = coefs[,
    "Estimate"], std.error = coefs[, "Std. Error"])
# Create a ggplot error bar chart
ebar_fe <- ggplot(coef_df, aes(x = term, y = estimate)) + geom_point() +
    coord_flip() + geom_hline(yintercept = 0, linetype = "dashed",
    color = "darkblue") + geom_errorbar(aes(ymin = estimate -
    std.error, ymax = estimate + std.error), width = 0.2) + labs(title = "Stage II: Static & Lagged TWFE
    x = NULL, y = "Coefficient Estimate") + theme_minimal() +
    # theme(axis.text.x = element_text(hjust = 1)) +
scale_x_discrete(labels = c(di_score = "Democracy (DI)", `plm::lag(di_score, 1)` = "Lagged DI (t-1)",
    `plm::lag(di_score, 2)` = "Lagged DI (t-2)", spi_comp = "Statistical Performance (SPI)",
    `plm::lag(spi_comp, 1)` = "Lagged SPI (t-1)", `plm::lag(spi_comp, 2)` = "Lagged SPI (t-2)",
    log_gdppc = "Log(GDP per capita)", `factor(year)2017` = "Year 2017",
    `factor(year)2018` = "Year 2018", `factor(year)2019` = "Year 2019",
    `factor(year)2020` = "Year 2020", `factor(year)2021` = "Year 2021",
    `factor(year)2022` = "Year 2022", `factor(year)2023` = "Year 2023",
    Intercept = "Intercept"), limits = c("di_score", "plm::lag(di_score, 1)",
    "plm::lag(di_score, 2)", "spi_comp", "plm::lag(spi_comp, 1)",
    "plm::lag(spi_comp, 2)", "log_gdppc", "factor(year)2017",
    "factor(year)2018", "factor(year)2019", "factor(year)2020",
    "factor(year)2021", "factor(year)2022", "factor(year)2023",
    "Intercept"))

ebar_fe
```

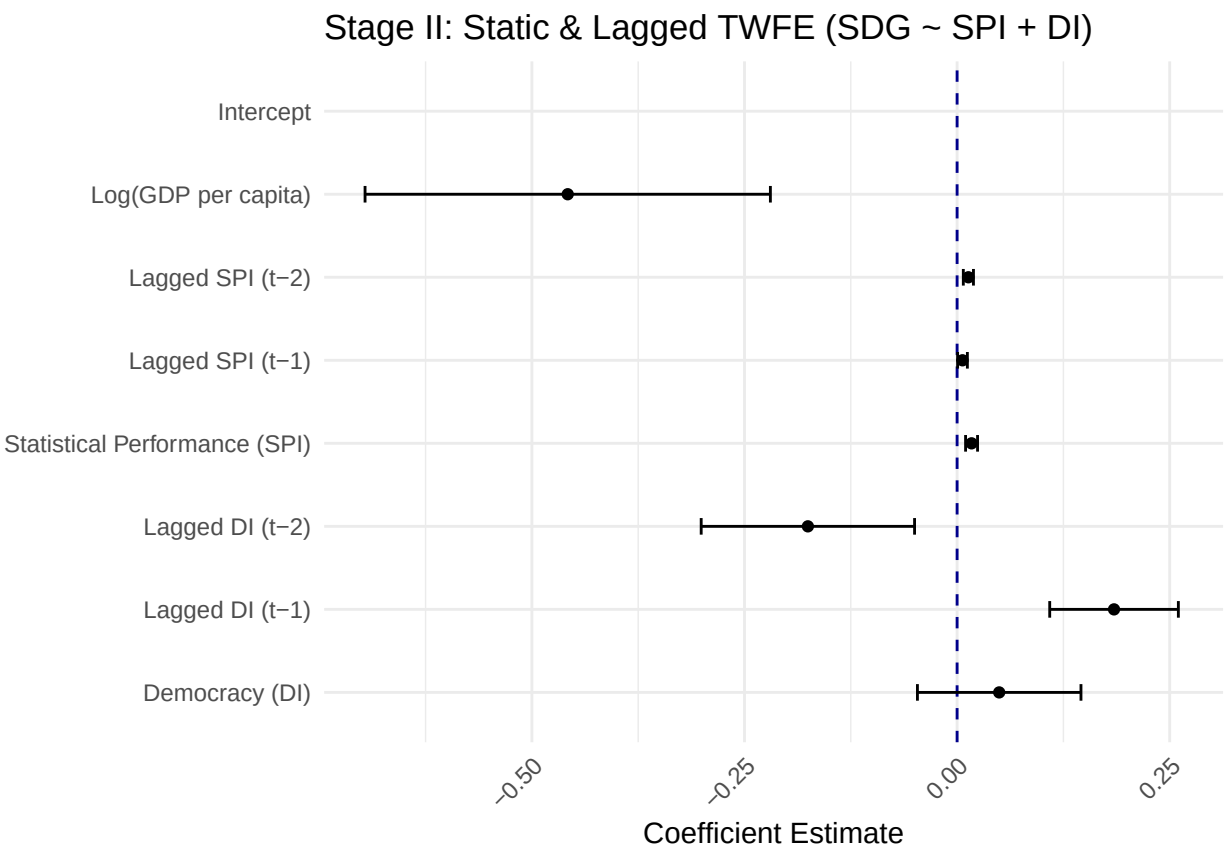


```r
# Save the plot
# ggsave('component_2/figures/stage2/error_bar_fe_sdg_spi_L2.png',
# ebar_fe, width = 10, height = 6)
```

5.3 First Difference Error Bar Visualization

```r
# Extract coefficients and robust standard errors from the
# FD model
coefs_fd <- summary(fd_sdg_spi_L2, vcov = vcovHC(fd_sdg_spi_L2,
    cluster = "group", type = "HC1"))$coefficients
# Create a data frame for visualization
coef_df_fd <- data.frame(term = rownames(coefs_fd), estimate = coefs_fd[,
    "Estimate"], std.error = coefs_fd[, "Std. Error"])
# Create a ggplot error bar chart for the FD model
ebar_fd <- ggplot(coef_df_fd, aes(x = term, y = estimate)) +
    geom_point() + coord_flip() + geom_hline(yintercept = 0,
    linetype = "dashed", color = "darkblue") + geom_errorbar(aes(ymin = estimate -
    std.error, ymax = estimate + std.error), width = 0.2) + labs(title = "Stage II: Static & Lagged TWFE
    x = NULL, y = "Coefficient Estimate") + theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
    scale_x_discrete(labels = c(di_score = "Democracy (DI)",
        `plm::lag(di_score, 1)` = "Lagged DI (t-1)", `plm::lag(di_score, 2)` = "Lagged DI (t-2)",
        spi_comp = "Statistical Performance (SPI)", `plm::lag(spi_comp, 1)` = "Lagged SPI (t-1)",
        `plm::lag(spi_comp, 2)` = "Lagged SPI (t-2)", log_gdppc = "Log(GDP per capita)",
        Intercept = "Intercept"), limits = c("di_score", "plm::lag(di_score, 1)",
        "plm::lag(di_score, 2)", "spi_comp", "plm::lag(spi_comp, 1)",
        "plm::lag(spi_comp, 2)", "log_gdppc", "Intercept"))

ebar_fd
```




```
# Save the plot
# ggsave('component_2/figures/stage2/error_bar_fd_sdg_spi_L2.png',
# ebar_fd, width = 10, height = 6)
```

5.4 Check for Autocorrelation

```
# APPLY Wooldridge Test for AR(1) Errors in FE Panel
# Models: pwartest()
# https://search.r-project.org/CRAN/refmans/plm/html/pwartest.html
# This is MUCH BETTER for panel data with small T AND
# unbalanced panels!!!
pwartest(fe_sdg_spi_L2)  # [significant]
```

```
##
##  Wooldridge's test for serial correlation in FE panels
##
## data:  fe_sdg_spi_L2
## F = 73.832, df1 = 1, df2 = 767, p-value < 2.2e-16
## alternative hypothesis: serial correlation
```

Significant p-value indicates the presence of autocorrelation in the residuals of the fixed effects model. This suggests that the errors are correlated over time, which violates one of the key assumptions of linear regression models.

5.5 Check for Heteroskedasticity

```
# Apply Breusch-Pagan test for heteroskedasticity
bptest(fe_sdg_spi_L2, studentize = TRUE)  # Heteroskedasticity [significant]
```

```
##
##  studentized Breusch-Pagan test
##
## data:  fe_sdg_spi_L2
## BP = 160.33, df = 12, p-value < 2.2e-16
```

```
bptest(fd_sdg_spi_L2, studentize = TRUE)  # Heteroskedasticity [significant]
```

```
##
##  studentized Breusch-Pagan test
##
## data:  fd_sdg_spi_L2
## BP = 148.67, df = 7, p-value < 2.2e-16
```

The Breusch-Pagan test indicates the presence of heteroskedasticity in the residuals of the fixed effects model. This suggests that the variance of the errors is not constant across observations, which violates another key assumption of linear regression models.

Both violations are corrected by using robust standard errors clustered at the country level, which accounts for autocorrelation and heteroskedasticity.