Component 2, Stage 2: SPI -> SDG

Sevastian Sanchez

2025-09-17

```
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_data <- panel_data %>%
    dplyr::select(country_name, country_code, year, sdg_overall,
        spi_comp, di_score, di_score_lag1, di_score_lag2, log_gdppc,
        spi_comp_lag1, spi_comp_lag2, income_level_recoded) %>%
    dplyr::arrange(country_code, year)

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

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

```
# check lag structure is correct
head(panel_data[, c("country_code", "year", "sdg_overall", "spi_comp",
    "spi_comp_lag1", "spi_comp_lag2", "di_score", "di_score_lag1",
    "di_score_lag2")])
```

```
## # A tibble: 6 x 9
##   country_code  year sdg_overall spi_comp spi_comp_lag1 spi_comp_lag2 di_score
##   <chr>        <dbl>       <dbl>    <dbl>         <dbl>         <dbl>    <dbl>
```

```
## 1 AFG           2016        44.3     39.4          NA           NA       2.55
## 2 AFG           2017        45.0     44.8          39.4         NA       2.55
## 3 AFG           2018        45.5     52.0          44.8         39.4     2.97
## 4 AFG           2019        46.2     51.9          52.0         44.8     2.85
## 5 AFG           2020        47.5     55.6          51.9         52.0     2.85
## 6 AFG           2021        46.2     59.2          55.6         51.9     0.32
## # i 2 more variables: di_score_lag1 <dbl>, di_score_lag2 <dbl>
```

```r
# check dimensions
dim(panel_data)
```

```
## [1] 1296   12
```

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", index = c("country_code", "year"), 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", index = c("country_code", "year"))
##
## 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 + spi_comp_lag1 +
    di_score + di_score_lag1 + log_gdppc + factor(income_level_recoded) +
    factor(year), model = "pooling", index = c("country_code",
    "year"), 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", type
##
## Call:
## plm(formula = sdg_overall ~ spi_comp + spi_comp_lag1 + di_score +
##     di_score_lag1 + log_gdppc + factor(income_level_recoded) +
##     factor(year), data = panel_data, model = "pooling", index = c("country_code",
##     "year"))
##
## Unbalanced Panel: n = 155, T = 5-7, N = 1079
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -11.688966  -2.748015   0.076454   2.788766  14.513012
##
## Coefficients:
##                                Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)                   27.736859   4.582090  6.0533 1.964e-09 ***
## spi_comp                       0.074196   0.058568  1.2668  0.205496
## spi_comp_lag1                  0.214180   0.053008  4.0405 5.717e-05 ***
## di_score                       0.897671   0.615320  1.4589  0.144896
## di_score_lag1                 -0.651101   0.636571 -1.0228  0.306623
## log_gdppc                      1.588840   0.742344  2.1403  0.032557 *
## factor(income_level_recoded)1  4.849109   1.236966  3.9202 9.415e-05 ***
## factor(income_level_recoded)2  8.484858   1.761058  4.8180 1.660e-06 ***
## factor(income_level_recoded)3  8.008641   2.641715  3.0316  0.002491 **
## factor(year)2018              -0.523275   0.158072 -3.3104  0.000963 ***
## factor(year)2019              -0.672302   0.242841 -2.7685  0.005730 **
## factor(year)2020              -0.400347   0.301988 -1.3257  0.185223
## factor(year)2021              -1.126009   0.431990 -2.6066  0.009273 **
## factor(year)2022              -2.177969   0.466997 -4.6638 3.499e-06 ***
## factor(year)2023              -2.179319   0.444974 -4.8976 1.120e-06 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    107440
## Residual Sum of Squares: 20538
## R-Squared:      0.80884
## Adj. R-Squared: 0.80632
## F-statistic: 73.4697 on 14 and 154 DF, p-value: < 2.22e-16
```

```r
# Adding Lag2: SPI ~ DI
ols_sdg_spi_L2 <- plm(formula = sdg_overall ~ spi_comp + spi_comp_lag1 +
    spi_comp_lag2 + di_score + di_score_lag1 + di_score_lag2 +
    log_gdppc + factor(income_level_recoded) + factor(year),
    model = "pooling", index = c("country_code", "year"), 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 + spi_comp_lag1 + spi_comp_lag2 +
##     di_score + di_score_lag1 + di_score_lag2 + log_gdppc + factor(income_level_recoded) +
##     factor(year), data = panel_data, model = "pooling", index = c("country_code",
##     "year"))
##
## Unbalanced Panel: n = 155, T = 4-6, N = 924
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -11.71018  -2.77075   0.13186   2.83273  14.00417
##
## Coefficients:
##                                 Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)                    27.587633   4.687507  5.8854 5.584e-09 ***
## spi_comp                        0.091007   0.065997  1.3790  0.168245
## spi_comp_lag1                   0.035694   0.032364  1.1029  0.270369
## spi_comp_lag2                   0.168055   0.054651  3.0751  0.002167 **
## di_score                        0.732140   0.557259  1.3138  0.189237
## di_score_lag1                   0.183004   0.342883  0.5337  0.593665
## di_score_lag2                  -0.722435   0.671343 -1.0761  0.282166
## log_gdppc                       1.556725   0.750747  2.0736  0.038401 *
## factor(income_level_recoded)1   5.032017   1.267522  3.9700 7.756e-05 ***
## factor(income_level_recoded)2   8.604467   1.793346  4.7980 1.873e-06 ***
## factor(income_level_recoded)3   8.189760   2.711183  3.0207  0.002592 **
## factor(year)2019               -0.099031   0.203239 -0.4873  0.626188
## factor(year)2020               -0.214918   0.274614 -0.7826  0.434055
## factor(year)2021               -0.626017   0.354438 -1.7662  0.077694 .
## factor(year)2022               -1.171814   0.383213 -3.0579  0.002295 **
## factor(year)2023               -2.186161   0.447007 -4.8907 1.188e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    91386
```

```
## Residual Sum of Squares: 17170
## R-Squared:      0.81212
## Adj. R-Squared: 0.80901
## F-statistic: 74.9353 on 15 and 154 DF, p-value: < 2.22e-16
```

2.1 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", index = c("country_code",
    "year"), 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 +
##     factor(income_level_recoded), data = fd_data, model = "fd",
##     index = c("country_code", "year"))
##
## Unbalanced Panel: n = 155, T = 6-8, N = 1234
## Observations used in estimation: 1079
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -1.934635 -0.315374 -0.046495  0.248608  3.136743
##
## Coefficients:
##                                 Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)                    0.3052187  0.0194557 15.6879 < 2.2e-16 ***
## spi_comp                       0.0161244  0.0060174  2.6797  0.007483 **
## di_score                       0.0391030  0.0844136  0.4632  0.643293
## log_gdppc                     -0.2403192  0.2205508 -1.0896  0.276120
## factor(income_level_recoded)1  0.0037130  0.1825353  0.0203  0.983775
## factor(income_level_recoded)2  0.1484906  0.2066353  0.7186  0.472537
## factor(income_level_recoded)3  0.1627284  0.2219296  0.7332  0.463570
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    348.2
## Residual Sum of Squares: 344.51
## R-Squared:      0.010603
## Adj. R-Squared: 0.0050657
## F-statistic: 1.72384 on 6 and 154 DF, p-value: 0.11891
```

```r
# Adding Lag1: SPI ~ DI
fd_sdg_spi_L1 <- plm(formula = sdg_overall ~ spi_comp + spi_comp_lag1 +
    di_score + di_score_lag1 + log_gdppc + factor(income_level_recoded),
    model = "fd", index = c("country_code", "year"), 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 + spi_comp_lag1 + di_score +
##     di_score_lag1 + log_gdppc + factor(income_level_recoded),
##     data = fd_data, model = "fd", index = c("country_code", "year"))
##
## Unbalanced Panel: n = 155, T = 5-7, N = 1079
## Observations used in estimation: 924
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -1.929824 -0.313471 -0.045799  0.252786  3.104474
##
## Coefficients:
##                                 Estimate Std. Error t-value Pr(>|t|)
## (Intercept)                    0.2853002  0.0250717 11.3794  < 2e-16 ***
## spi_comp                       0.0115543  0.0064691  1.7861  0.07442 .
## spi_comp_lag1                  0.0079235  0.0060511  1.3094  0.19071
## di_score                       0.0727949  0.0843416  0.8631  0.38831
## di_score_lag1                  0.1349714  0.0744059  1.8140  0.07001 .
## log_gdppc                     -0.3505928  0.2483091 -1.4119  0.15831
## factor(income_level_recoded)1  0.0461826  0.2099446  0.2200  0.82594
## factor(income_level_recoded)2  0.2726074  0.2367688  1.1514  0.24988
## factor(income_level_recoded)3  0.3359719  0.2528855  1.3286  0.18433
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    303.64
## Residual Sum of Squares: 299.2
## R-Squared:      0.014622
## Adj. R-Squared: 0.006007
## F-statistic: 1.67189 on 8 and 154 DF, p-value: 0.10937
```

```r
# Adding Lag2: SPI ~ DI
fd_sdg_spi_L2 <- plm(formula = sdg_overall ~ spi_comp + spi_comp_lag1 +
    spi_comp_lag2 + di_score + di_score_lag1 + di_score_lag2 +
    log_gdppc + factor(income_level_recoded), model = "fd", index = c("country_code",
    "year"), 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 + spi_comp_lag1 + spi_comp_lag2 +
##     di_score + di_score_lag1 + di_score_lag2 + log_gdppc + factor(income_level_recoded),
##     data = fd_data, model = "fd", index = c("country_code", "year"))
##
## Unbalanced Panel: n = 155, T = 4-6, N = 924
## Observations used in estimation: 769
##
## Residuals:
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -1.85996 -0.30202 -0.03658  0.25639  2.91190
##
## Coefficients:
##                                  Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)                     0.2339583  0.0283691  8.2469 7.159e-16 ***
## spi_comp                        0.0175739  0.0070274  2.5008   0.01260 *
## spi_comp_lag1                   0.0069870  0.0057285  1.2197   0.22296
## spi_comp_lag2                   0.0133215  0.0059503  2.2388   0.02546 *
## di_score                        0.0575724  0.0975677  0.5901   0.55532
## di_score_lag1                   0.1834978  0.0765475  2.3972   0.01676 *
## di_score_lag2                  -0.1765615  0.1262990 -1.3980   0.16253
## log_gdppc                      -0.4945605  0.2413981 -2.0487   0.04083 *
## factor(income_level_recoded)1  0.1177717  0.1735894  0.6784   0.49769
## factor(income_level_recoded)2  0.3190698  0.2020118  1.5795   0.11465
## factor(income_level_recoded)3  0.3913713  0.2137640  1.8309   0.06751 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    229.74
## Residual Sum of Squares: 221.66
## R-Squared:      0.035145
## Adj. R-Squared: 0.022416
## F-statistic: 2.16212 on 10 and 154 DF, p-value: 0.022917
```

3.1 Stargazer Table for FD Models

4 2.3) Fixed Effects [Stage 2]

```r
# Contemporaneous Effect: SDG ~ SPI + DI
fe_sdg_spi <- plm(formula = sdg_overall ~ spi_comp + di_score +
    log_gdppc + factor(year) + factor(income_level_recoded),
    index = c("country_code", "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) + factor(income_level_recoded), data = panel_data,
##     model = "within", index = c("country_code", "year"))
##
## Unbalanced Panel: n = 155, T = 6-8, N = 1234
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -2.8369120 -0.3092824 -0.0081391  0.3334427  3.4434607
##
## Coefficients:
##                                 Estimate Std. Error t-value  Pr(>|t|)
## spi_comp                        0.044006   0.014857  2.9619  0.003126 **
## di_score                        0.117373   0.089489  1.3116  0.189938
## log_gdppc                       0.337323   0.315447  1.0693  0.285154
## factor(year)2017                0.342538   0.053932  6.3512 3.157e-10 ***
## factor(year)2018                0.554374   0.092605  5.9865 2.927e-09 ***
## factor(year)2019                0.946068   0.098688  9.5865 < 2.2e-16 ***
## factor(year)2020                1.258547   0.108831 11.5642 < 2.2e-16 ***
## factor(year)2021                1.248780   0.164380  7.5969 6.617e-14 ***
## factor(year)2022                1.445105   0.166247  8.6925 < 2.2e-16 ***
## factor(year)2023                1.616792   0.182979  8.8359 < 2.2e-16 ***
## factor(income_level_recoded)1 0.391704   0.398058  0.9840  0.325320
## factor(income_level_recoded)2 0.738052   0.475931  1.5508  0.121258
## factor(income_level_recoded)3 0.371058   0.545743  0.6799  0.496707
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    1162.9
## Residual Sum of Squares: 441.55
## R-Squared:      0.6203
## Adj. R-Squared: 0.56082
## F-statistic: 45.9504 on 13 and 154 DF, p-value: < 2.22e-16
```

```
# Adding Lag1: SPI ~ DI
fe_sdg_spi_L1 <- plm(formula = sdg_overall ~ spi_comp + spi_comp_lag1 +
    di_score + di_score_lag1 + log_gdppc + factor(income_level_recoded) +
    factor(year), model = "within", index = c("country_code",
    "year"), 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 + spi_comp_lag1 + di_score +
##     di_score_lag1 + log_gdppc + factor(income_level_recoded) +
##     factor(year), data = panel_data, model = "within", index = c("country_code",
##     "year"))
##
```

```
## Unbalanced Panel: n = 155, T = 5-7, N = 1079
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -2.658946 -0.271198 -0.013632  0.271782  3.140399
##
## Coefficients:
##                                 Estimate Std. Error t-value  Pr(>|t|)
## spi_comp                       0.0241741  0.0112967  2.1399   0.03263 *
## spi_comp_lag1                  0.0237987  0.0096619  2.4632   0.01396 *
## di_score                       0.2374865  0.0942257  2.5204   0.01189 *
## di_score_lag1                 -0.0458757  0.1096028 -0.4186   0.67563
## log_gdppc                      0.3339167  0.3676600  0.9082   0.36400
## factor(income_level_recoded)1  0.3728521  0.3144906  1.1856   0.23610
## factor(income_level_recoded)2  0.7585034  0.4086083  1.8563   0.06373 .
## factor(income_level_recoded)3  0.3845908  0.5307062  0.7247   0.46884
## factor(year)2018               0.2049135  0.0641109  3.1962   0.00144 **
## factor(year)2019               0.5479565  0.0781320  7.0132 4.541e-12 ***
## factor(year)2020               0.9074314  0.0921364  9.8488 < 2.2e-16 ***
## factor(year)2021               0.9466398  0.1360859  6.9562 6.676e-12 ***
## factor(year)2022               1.0123227  0.1617134  6.2600 5.925e-10 ***
## factor(year)2023               1.2283736  0.1690011  7.2684 7.826e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    740.6
## Residual Sum of Squares: 312.19
## R-Squared:      0.57847
## Adj. R-Squared: 0.50065
## F-statistic: 33.8593 on 14 and 154 DF, p-value: < 2.22e-16
```

```
# Adding Lag2: SPI ~ DI
fe_sdg_spi_L2 <- plm(formula = sdg_overall ~ spi_comp + spi_comp_lag1 +
    spi_comp_lag2 + di_score + di_score_lag1 + di_score_lag2 +
    log_gdppc + factor(income_level_recoded) + factor(year),
    model = "within", index = c("country_code", "year"), data = panel_data)
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 + spi_comp_lag1 + spi_comp_lag2 +
##     di_score + di_score_lag1 + di_score_lag2 + log_gdppc + factor(income_level_recoded) +
##     factor(year), data = panel_data, model = "within", index = c("country_code",
##     "year"))
##
## Unbalanced Panel: n = 155, T = 4-6, N = 924
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -2.4105174 -0.2462874 -0.0016371  0.2405397  2.8719407
```

```
##
## Coefficients:
##                                  Estimate Std. Error t-value  Pr(>|t|)
## spi_comp                         0.0338896  0.0117720  2.8788  0.004105 **
## spi_comp_lag1                    0.0063699  0.0075753  0.8409  0.400685
## spi_comp_lag2                    0.0163276  0.0085846  1.9020  0.057557 .
## di_score                         0.1559032  0.0825031  1.8897  0.059186 .
## di_score_lag1                    0.1701864  0.0971166  1.7524  0.080113 .
## di_score_lag2                   -0.2460409  0.1258641 -1.9548  0.050974 .
## log_gdppc                       -0.0725105  0.3338289 -0.2172  0.828105
## factor(income_level_recoded)1    0.4185389  0.1920359  2.1795  0.029605 *
## factor(income_level_recoded)2    0.7948941  0.2748999  2.8916  0.003943 **
## factor(income_level_recoded)3    0.5141073  0.3689604  1.3934  0.163911
## factor(year)2019                 0.3334744  0.0438533  7.6043 8.534e-14 ***
## factor(year)2020                 0.6107261  0.0763721  7.9967 4.803e-15 ***
## factor(year)2021                 0.6878165  0.1068534  6.4370 2.167e-10 ***
## factor(year)2022                 0.8321507  0.1292528  6.4382 2.151e-10 ***
## factor(year)2023                 0.9354750  0.1628894  5.7430 1.349e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    440.52
## Residual Sum of Squares: 198.57
## R-Squared:      0.54924
## Adj. R-Squared: 0.44821
## F-statistic: 29.6661 on 15 and 154 DF, p-value: < 2.22e-16
```

5 Stargazer Table for FE Models

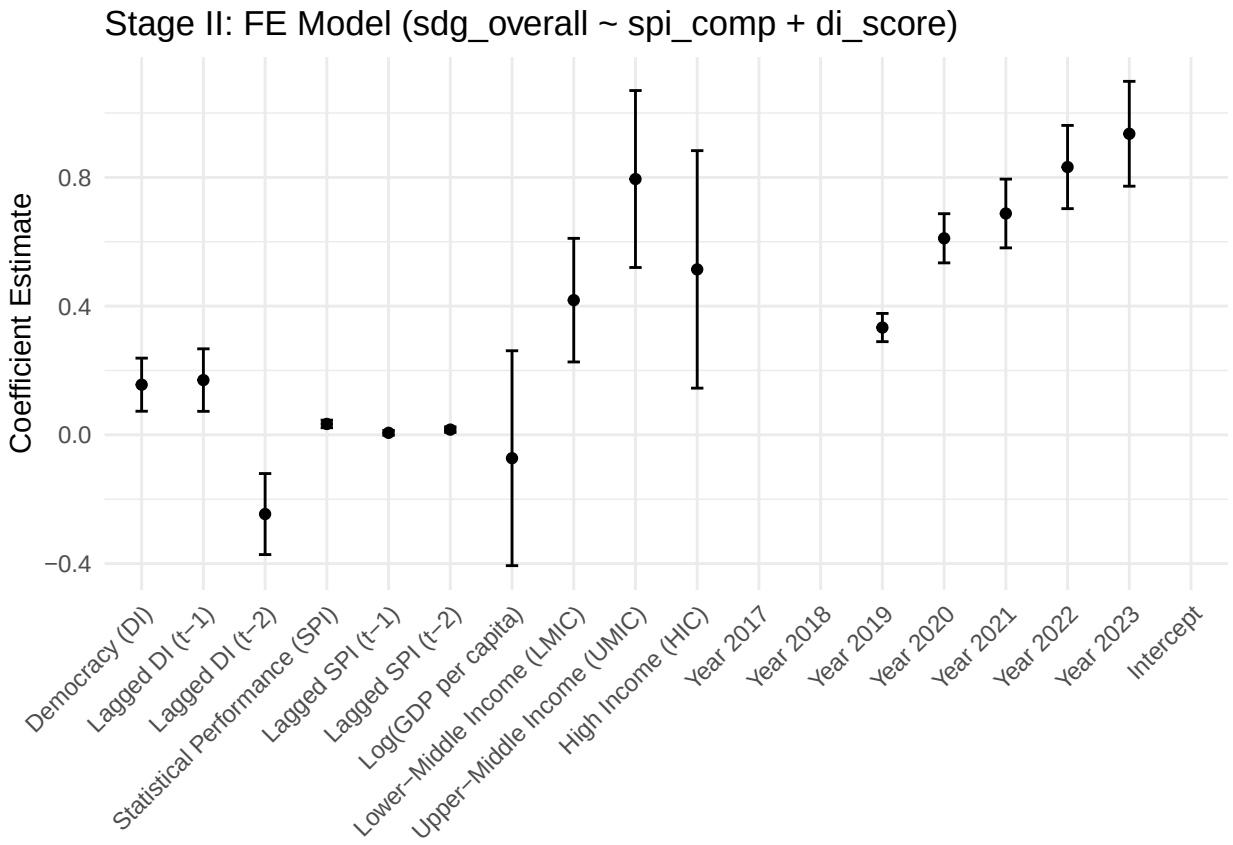
5.1 make a stargazer table of all Lag2 models for POLS, FD and FE

5.2 Fixed Effects Error Bar Visualization

```
# Extract coefficients and robust standard errors from the
# FE model
coefs <- summary(fe_sdg_spi_L2, vcov = vcovHC(fe_sdg_spi_L2,
    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() +
    geom_errorbar(aes(ymin = estimate - std.error, ymax = estimate +
        std.error), width = 0.2) + labs(title = "Stage II: FE Model (sdg_overall ~ spi_comp + di_score)
    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)",
        di_score_lag1 = "Lagged DI (t-1)", di_score_lag2 = "Lagged DI (t-2)",
        spi_comp = "Statistical Performance (SPI)", spi_comp_lag1 = "Lagged SPI (t-1)",
        spi_comp_lag2 = "Lagged SPI (t-2)", log_gdppc = "Log(GDP per capita)",
        `factor(income_level_recoded)1` = "Lower-Middle Income (LMIC)",
```

```
        `factor(income_level_recoded)2` = "Upper-Middle Income (UMIC)",
        `factor(income_level_recoded)3` = "High Income (HIC)",
        `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", "di_score_lag1", "di_score_lag2",
            "spi_comp", "spi_comp_lag1", "spi_comp_lag2", "log_gdppc",
            "factor(income_level_recoded)1", "factor(income_level_recoded)2",
            "factor(income_level_recoded)3", "factor(year)2017",
            "factor(year)2018", "factor(year)2019", "factor(year)2020",
            "factor(year)2021", "factor(year)2022", "factor(year)2023",
            "Intercept"))

ebar_fe
```



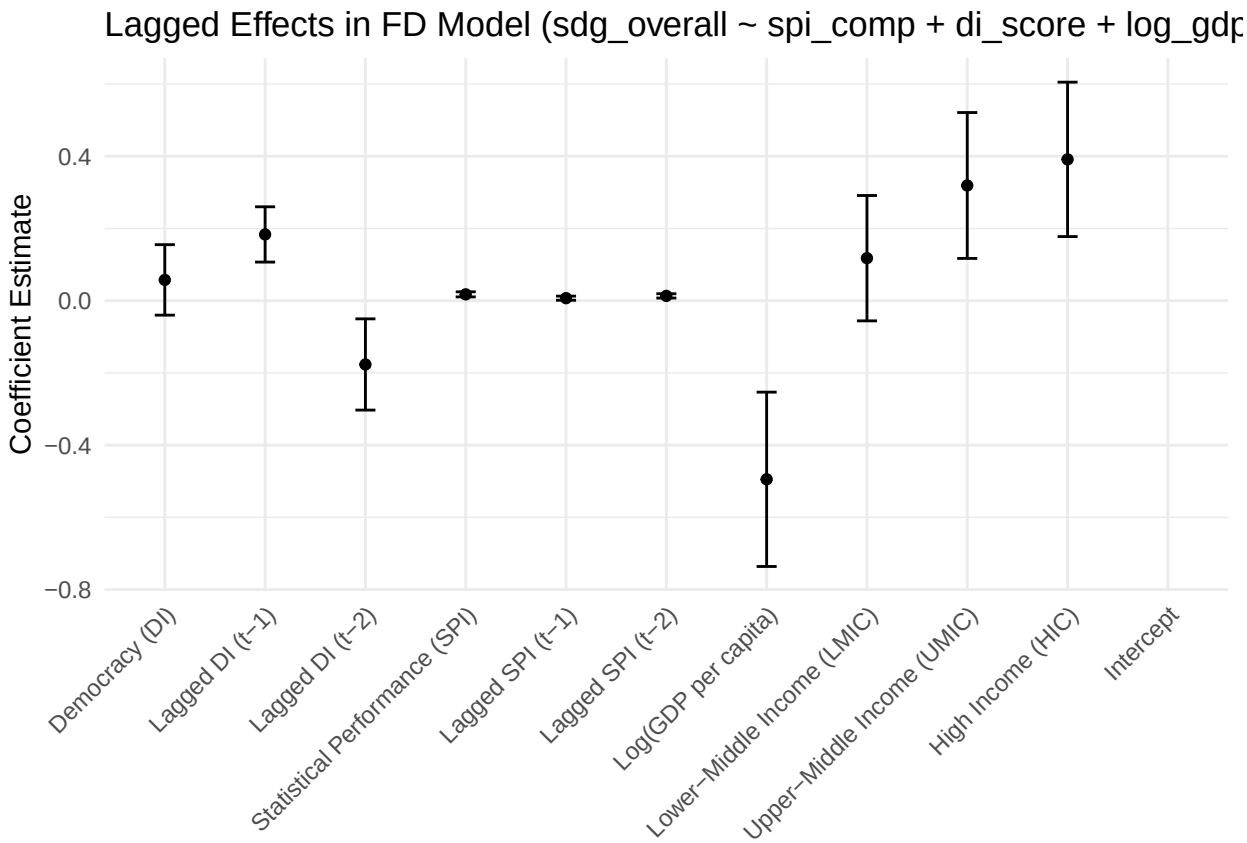

```
# Save the plot
# ggsave('component_2/figures/stage2/error_bar_fe_sdg_spi_L2.png',
# ebar_fe, width = 10, height = 6)
```

5.3 First Difference Error Bar Visualization

```
# Extract coefficients and robust standard errors from the
# FD model
```

```r
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() + geom_errorbar(aes(ymin = estimate - std.error,
    ymax = estimate + std.error), width = 0.2) + labs(title = "Lagged Effects in FD Model (sdg_overall
    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)",
        di_score_lag1 = "Lagged DI (t-1)", di_score_lag2 = "Lagged DI (t-2)",
        spi_comp = "Statistical Performance (SPI)", spi_comp_lag1 = "Lagged SPI (t-1)",
        spi_comp_lag2 = "Lagged SPI (t-2)", log_gdppc = "Log(GDP per capita)",
        `factor(income_level_recoded)1` = "Lower-Middle Income (LMIC)",
        `factor(income_level_recoded)2` = "Upper-Middle Income (UMIC)",
        `factor(income_level_recoded)3` = "High Income (HIC)",
        Intercept = "Intercept"), limits = c("di_score", "di_score_lag1",
        "di_score_lag2", "spi_comp", "spi_comp_lag1", "spi_comp_lag2",
        "log_gdppc", "factor(income_level_recoded)1", "factor(income_level_recoded)2",
        "factor(income_level_recoded)3", "Intercept"))

ebar_fd
```



```r
# 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

```r
# 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 = 67.496, df1 = 1, df2 = 767, p-value = 8.955e-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

```r
# 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 = 98.322, df = 15, p-value = 2.711e-14
```

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

```
##
##  studentized Breusch-Pagan test
##
## data:  fd_sdg_spi_L2
## BP = 93.383, df = 10, p-value = 1.14e-15
```

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.