Component 2, Stage 2: SPI -> SDG

Sevastian Sanchez

2025-09-02

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

Set up

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

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

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.4
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
## Loading required package: carData
##
##
## Attaching package: 'car'
##
##
## The following object is masked from 'package:dplyr':
##
##     recode
##
##
## The following object is masked from 'package:purrr':
##
##     some
##
##
## Loading required package: usethis
##
##
```

```
## Attaching package: 'ERT'
##
##
## The following objects are masked from 'package:vdemdata':
##
##     codebook, vdem
##
##
##
## Please cite as:
##
##
##  Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
##
##  R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
##
##
##
## Attaching package: 'scales'
##
##
## The following object is masked from 'package:purrr':
##
##     discard
##
##
## The following object is masked from 'package:readr':
##
##     col_factor
##
##
##
## Attaching package: 'kableExtra'
##
##
## The following object is masked from 'package:dplyr':
##
##     group_rows
##
##
##
## Attaching package: 'mice'
##
##
## The following object is masked from 'package:stats':
##
##     filter
##
##
## The following objects are masked from 'package:base':
##
##     cbind, rbind
##
##
```

```
## Loading required package: MASS
##
##
## Attaching package: 'MASS'
##
##
## The following object is masked from 'package:dplyr':
##
##     select
##
##
##
## Attaching package: 'plm'
##
##
## The following objects are masked from 'package:dplyr':
##
##     between, lag, lead
##
##
##
## Attaching package: 'patchwork'
##
##
## The following object is masked from 'package:MASS':
##
##     area
##
##
##
## Attaching package: 'reshape2'
##
##
## The following object is masked from 'package:tidyr':
##
##     smiths
##
##
##
## Attaching package: 'jsonlite'
##
##
## The following object is masked from 'package:purrr':
##
##     flatten
##
##
## Loading required package: zoo
##
##
## Attaching package: 'zoo'
##
##
## The following objects are masked from 'package:base':
```

```
##
##     as.Date, as.Date.numeric
##
##
## Loading required package: Matrix
##
##
## Attaching package: 'Matrix'
##
##
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
##
##
## Loading required package: mvtnorm
##
## mediation: Causal Mediation Analysis
## Version: 4.5.0
##
##
##
## Attaching package: 'plotly'
##
##
## The following object is masked from 'package:MASS':
##
##     select
##
##
## The following object is masked from 'package:ggplot2':
##
##     last_plot
##
##
## The following object is masked from 'package:stats':
##
##     filter
##
##
## The following object is masked from 'package:graphics':
##
##     layout
##
##
##
## Attaching package: 'ggdag'
##
##
## The following object is masked from 'package:stats':
##
##     filter
```

```r
# load data
source("Comp2_panel_wrangling.R")
```

```
## Rows: 3340 Columns: 70
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (6): country_name, country_code, income_level, income_spi, region_spi, ...
## dbl (64): year, year_fct, sdg_overall, spi_comp, sci_overall, di_score, di_r...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 179 Columns: 6
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): country_code, country_name
## dbl (4): in_merged_cleaned_spi, in_merged_cleaned_sdg, in_merged_exclusive, ...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
# select path = 'data/Main CSV
# Outputs/merged_cleaned_sdg.csv'

# select relevant variables and arrange data
panel_data_s2 <- 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)

# check lag structure is correct
head(panel_data_s2[, 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
dim(panel_data_s2)  # 1336 rows, 12 columns
```

```
## [1] 1336   12
```

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_s2)
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_s2,
##     model = "pooling", index = c("country_code", "year"))
##
## Unbalanced Panel: n = 156, T = 2-8, N = 1236
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -11.97968  -2.87083   0.10677   2.84991  14.07621
##
## Coefficients:
##                                Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)                    27.31694    4.48773  6.0870 1.537e-09 ***
## spi_comp                        0.28502    0.03396  8.3928 < 2.2e-16 ***
## di_score                        0.29521    0.25569  1.1546 0.2484983
## log_gdppc                       1.61290    0.73529  2.1936 0.0284553 *
## factor(income_level_recoded)1   4.75104    1.23474  3.8478 0.0001253 ***
## factor(income_level_recoded)2   8.49446    1.74225  4.8756 1.228e-06 ***
## factor(income_level_recoded)3   7.89851    2.58014  3.0613 0.0022521 **
## factor(year)2017               -0.27618    0.15416 -1.7915 0.0734626 .
## factor(year)2018               -0.83293    0.25212 -3.3037 0.0009819 ***
## factor(year)2019               -0.53331    0.27157 -1.9638 0.0497804 *
## factor(year)2020               -0.71328    0.36377 -1.9608 0.0501303 .
## factor(year)2021               -1.96984    0.49639 -3.9684 7.658e-05 ***
## factor(year)2022               -1.84644    0.47975 -3.8487 0.0001249 ***
## factor(year)2023               -2.26316    0.49228 -4.5972 4.725e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    125210
## Residual Sum of Squares: 24054
## R-Squared:      0.8079
## Adj. R-Squared: 0.80585
## F-statistic: 76.0877 on 13 and 155 DF, p-value: < 2.22e-16
```

```
# 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_s2)
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_s2, model = "pooling", index = c("country_code",
##     "year"))
##
## Unbalanced Panel: n = 156, T = 1-7, N = 1080
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -11.70142  -2.75172   0.08296   2.79950  14.54241
##
## Coefficients:
##                                 Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)                    27.676956   4.576734  6.0473 2.036e-09 ***
## spi_comp                        0.072523   0.058586  1.2379 0.2160308
## spi_comp_lag1                   0.215984   0.052947  4.0792 4.856e-05 ***
## di_score                        0.903338   0.615170  1.4684 0.1422810
## di_score_lag1                  -0.655383   0.636700 -1.0293 0.3035519
## log_gdppc                       1.593959   0.742400  2.1470 0.0320157 *
## factor(income_level_recoded)1   4.864444   1.235806  3.9363 8.814e-05 ***
## factor(income_level_recoded)2   8.492487   1.761244  4.8219 1.629e-06 ***
## factor(income_level_recoded)3   8.003483   2.642156  3.0291 0.0025116 **
## factor(year)2018               -0.524266   0.158169 -3.3146 0.0009487 ***
## factor(year)2019               -0.677143   0.242669 -2.7904 0.0053584 **
## factor(year)2020               -0.401666   0.302199 -1.3291 0.1840850
## factor(year)2021               -1.123615   0.432478 -2.5981 0.0095039 **
## factor(year)2022               -2.186005   0.466632 -4.6846 3.167e-06 ***
## factor(year)2023               -2.209026   0.444319 -4.9717 7.730e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    107980
## Residual Sum of Squares: 20553
## R-Squared:      0.80965
## Adj. R-Squared: 0.80715
## F-statistic: 74.8231 on 14 and 155 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_s2)
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", 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_s2, 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
```

Stargazer Table for POLS Models

2.2) First Difference [Stage 2]

```r
# 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 = 156, T = 2-8, N = 1236
## Observations used in estimation: 1080
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -1.935895 -0.316227 -0.046961  0.249059  3.135173
##
## Coefficients:
##                                Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)                   0.3055624  0.0194511 15.7093 < 2.2e-16 ***
## spi_comp                      0.0165248  0.0060148  2.7474  0.006108 **
## di_score                      0.0363020  0.0845982  0.4291  0.667928
## log_gdppc                    -0.2387810  0.2205717 -1.0826  0.279249
## factor(income_level_recoded)1  0.0027935  0.1825908  0.0153  0.987796
## factor(income_level_recoded)2  0.1471402  0.2066928  0.7119  0.476695
## factor(income_level_recoded)3  0.1610575  0.2219793  0.7256  0.468272
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    350.18
## Residual Sum of Squares: 346.36
## R-Squared:      0.010901
## Adj. R-Squared: 0.0053697
## F-statistic: 1.77333 on 6 and 155 DF, p-value: 0.10796
```

```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 = 156, T = 1-7, N = 1080
## 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  0.0250600 11.3847  < 2e-16 ***
## spi_comp                     0.0115543  0.0064660  1.7869  0.07428 .
```

```
## spi_comp_lag1                  0.0079235  0.0060482  1.3101  0.19050
## di_score                       0.0727949  0.0843022  0.8635  0.38809
## di_score_lag1                  0.1349714  0.0743712  1.8148  0.06988 .
## log_gdppc                     -0.3505928  0.2481932 -1.4126  0.15812
## factor(income_level_recoded)1  0.0461826  0.2098466  0.2201  0.82586
## factor(income_level_recoded)2  0.2726074  0.2366583  1.1519  0.24966
## factor(income_level_recoded)3  0.3359719  0.2527674  1.3292  0.18412
## ---
## 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: 1.67346 on 8 and 155 DF, p-value: 0.10892
```

```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
```

Stargazer Table for FD Models

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_s2,
    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_s2,
##     model = "within", index = c("country_code", "year"))
##
## Unbalanced Panel: n = 156, T = 2-8, N = 1236
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -2.8415475 -0.3119499 -0.0080513  0.3344781  3.4361363
##
## Coefficients:
##                                  Estimate Std. Error t-value  Pr(>|t|)
## spi_comp                         0.044217   0.014835  2.9806  0.002942 **
## di_score                         0.116711   0.089502  1.3040  0.192512
## log_gdppc                        0.338504   0.315634  1.0725  0.283757
## factor(year)2017                 0.341924   0.053906  6.3430 3.322e-10 ***
## factor(year)2018                 0.553193   0.092549  5.9773 3.089e-09 ***
## factor(year)2019                 0.944832   0.098648  9.5779 < 2.2e-16 ***
## factor(year)2020                 1.256826   0.108732 11.5590 < 2.2e-16 ***
## factor(year)2021                 1.245846   0.164175  7.5885 7.032e-14 ***
## factor(year)2022                 1.437982   0.166018  8.6616 < 2.2e-16 ***
## factor(year)2023                 1.617979   0.182909  8.8458 < 2.2e-16 ***
## factor(income_level_recoded)1 0.391069   0.398032  0.9825  0.326073
## factor(income_level_recoded)2 0.736325   0.475928  1.5471  0.122127
## factor(income_level_recoded)3 0.367749   0.545828  0.6737  0.500619
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    1164.4
```

```
## Residual Sum of Squares: 442.42
## R-Squared:      0.62004
## Adj. R-Squared: 0.56021
## F-statistic: 45.7442 on 13 and 155 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_s2)
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_s2, model = "within", index = c("country_code",
##     "year"))
##
## Unbalanced Panel: n = 156, T = 1-7, N = 1080
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -2.658946 -0.271073 -0.013536  0.271602  3.140399
##
## Coefficients:
##                                  Estimate Std. Error t-value  Pr(>|t|)
## spi_comp                        0.0241741  0.0112966  2.1399   0.03262 *
## spi_comp_lag1                   0.0237987  0.0096618  2.4632   0.01396 *
## di_score                        0.2374865  0.0942251  2.5204   0.01189 *
## di_score_lag1                  -0.0458757  0.1096022 -0.4186   0.67563
## log_gdppc                       0.3339167  0.3676577  0.9082   0.36400
## factor(income_level_recoded)1   0.3728521  0.3144887  1.1856   0.23610
## factor(income_level_recoded)2   0.7585034  0.4086058  1.8563   0.06373 .
## factor(income_level_recoded)3   0.3845908  0.5307029  0.7247   0.46883
## factor(year)2018                0.2049135  0.0641105  3.1963   0.00144 **
## factor(year)2019                0.5479565  0.0781315  7.0133 4.540e-12 ***
## factor(year)2020                0.9074314  0.0921359  9.8488 < 2.2e-16 ***
## factor(year)2021                0.9466398  0.1360851  6.9562 6.675e-12 ***
## factor(year)2022                1.0123227  0.1617124  6.2600 5.923e-10 ***
## factor(year)2023                1.2283736  0.1690001  7.2685 7.823e-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: 0.50018
## F-statistic: 33.8597 on 14 and 155 DF, p-value: < 2.22e-16
```

```r
# 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_s2)
summary(fe_sdg_spi_L2, vcov = vcovHC(fe_sdg_spi_L2, cluster = "group",
    type = "HC1"))
```

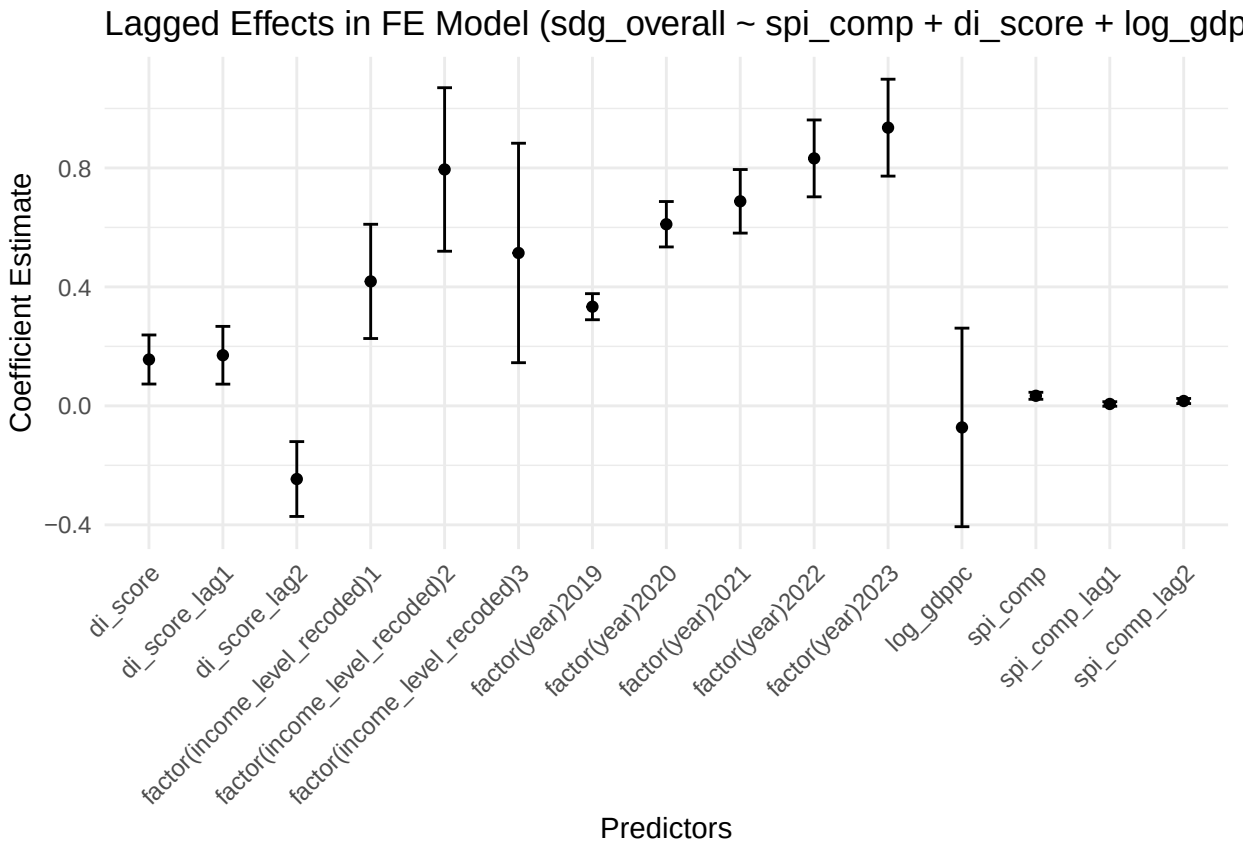
```
## Oneway (individual) effect Within Model
##
## Note: Coefficient variance-covariance matrix supplied: vcovHC(fe_sdg_spi_L2, cluster = "group", type
##
## Call:
## plm(formula = sdg_overall ~ spi_comp + spi_comp_lag1 + spi_comp_lag2 +
##     di_score + di_score_lag1 + di_score_lag2 + log_gdppc + factor(income_level_recoded) +
##     factor(year), data = panel_data_s2, 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
```

Stargazer Table for FE Models

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

Fixed Effects Error Bar Visualization

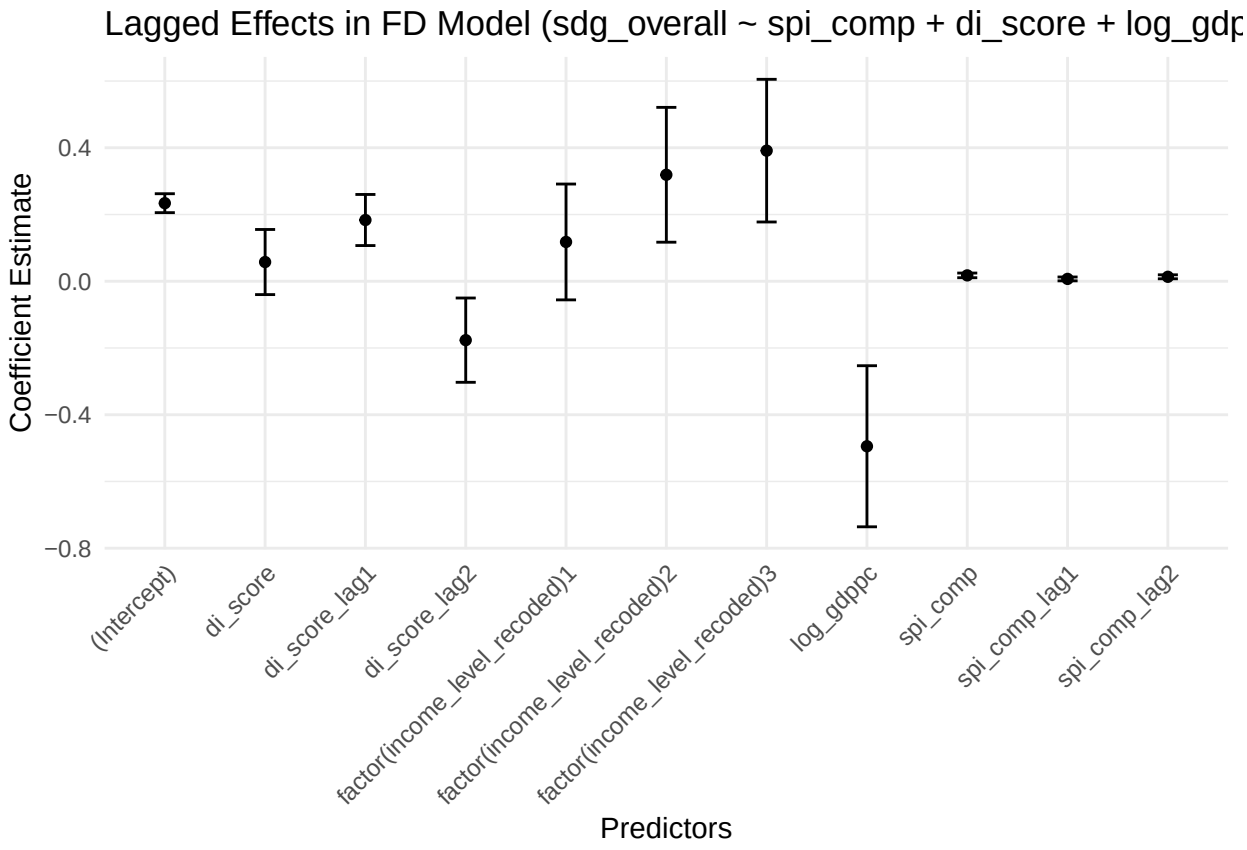
```
# Extract coefficients and robust standard errors from the
# FE model
coefs <- summary(fe_sdg_spi_L2, vcov = vcovHC(fe_sdg_spi_L2,
    cluster = "group", type = "HC1"))$coefficients
# Create a data frame for visualization
coef_df <- data.frame(term = rownames(coefs), estimate = coefs[,
    "Estimate"], std.error = coefs[, "Std. Error"])
# Create a ggplot error bar chart
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 = "Lagged Effects in FE Model (sdg_overall ~ spi_comp + di
    x = "Predictors", y = "Coefficient Estimate") + theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




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

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
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 ~ spi_comp + di
    x = "Predictors", y = "Coefficient Estimate") + theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




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

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.

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

```
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.

##[IGNORE ALL BELOW] # MODEL MISSPECIFICATION CHECKS [Stage 2] ========== ## RESET Test for Misspecification [STAGE 2] [ALL VARIABLES NOT UPDATED TO MATCH PRE-MADE LAGS]

Results from the reset test indicate that the null hypothesis of is rejected for all models, suggesting potential non-linearity or omitted variable bias in the models.

Stepwise Check: Applying Polynomial Terms [Stage 2]

For this section, all continuous predictors (DI, SPI, Log(GDP)) in non-linear models are centered to avoid multicollinearity issues.

- **H0:** The relationship between SPI and SDG performance is linear.
- **H1:** The relationship between SPI and SDG performance is non-linear (quadratic or cubic).

Model Selection: Adj. R^2 & AIC/BIC [Stage 2]

M1: standard FE AIC: 2324.805 BIC: 2447.637 #lowest BIC

M2: quadratic SPI, FE AIC: 2307.741 #lowest AIC BIC: 2451.045

M3: quadratic SPI + DI, FE AIC: 2313.891 BIC: 2477.667

M4: cubic SPI + DI, FE AIC: 2324.069 BIC: 2528.789