Component 2, Stage 1: DI -> SPI

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'

panel_data <- panel_data %>%
    dplyr::select(country_name, country_code, year, 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 that lag structure is correct
head(panel_data[, c("country_code", "year", "spi_comp", "spi_comp_lag1",
    "spi_comp_lag2", "di_score", "di_score_lag1", "di_score_lag2")])
```

```
## # A tibble: 6 x 8
##   country_code  year spi_comp spi_comp_lag1 spi_comp_lag2 di_score di_score_lag1
##   <chr>        <dbl>    <dbl>         <dbl>         <dbl>    <dbl>         <dbl>
## 1 AFG           2016     39.4          NA            NA       2.55          2.77
## 2 AFG           2017     44.8          39.4          NA       2.55          2.55
```

```
## 3 AFG           2018     52.0          44.8          39.4     2.97          2.55
## 4 AFG           2019     51.9          52.0          44.8     2.85          2.97
## 5 AFG           2020     55.6          51.9          52.0     2.85          2.85
## 6 AFG           2021     59.2          55.6          51.9     0.32          2.85
## # i 1 more variable: di_score_lag2 <dbl>
```

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

```
## [1] 1296   11
```

```r
# testing dataframes for sensitivity of results panel_data
# <- panel_data_spi
```

sensitivity analysis Basing analysis of dataset sensitivity on FE model of the second order (fe_spi_di_L2)

- Dataset = panel_data_comp1_data, 1336 obs, 167 countries, FE models (di_score_lag2) showed p = 0.07598 (marginally significant)
- Dataset = panel_data_sdg, 1336 obs, 167 countries, FE models (di_score_lag2) showed p = 0.07598 (marginally significant)
- Dataset = panel_data_exclusive, 1296 obs, 162 countries, FE models (di_score_lag2) showed p = 0.079 (marginally significant)
- Dataset = panel_data_spi, 1392 obs, 174 countries, FE models (di_score_lag2) showed p = 0.1497 (not significant)

2 Stage 1 Models:

ols_spi_di ols_spi_di_L1 ols_spi_di_L2

fd_spi_di fd_spi_di_L1 fd_spi_di_L2

fe_spi_di fe_spi_di_L1 fe_spi_di_L2

3 1.1) POLS [Stage 1]: SPI ~ DI

The effect of democracy on SPI Performance

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

```
## Pooling Model
##
## Note: Coefficient variance-covariance matrix supplied: vcovHC(ols_spi_di, cluster = "group", type =
```

```
##
## Call:
## plm(formula = 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.
## -33.45560  -5.38601   0.85982   6.43616  27.29382
##
## Coefficients:
##                                Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)                    15.38000    9.85158  1.5612   0.11874
## di_score                        3.57341    0.55240  6.4689 1.425e-10 ***
## log_gdppc                       2.69607    1.61716  1.6672   0.09574 .
## factor(income_level_recoded)1   1.24718    2.60136  0.4794   0.63172
## factor(income_level_recoded)2   1.84883    4.20052  0.4401   0.65991
## factor(income_level_recoded)3   3.28757    6.36887  0.5162   0.60581
## factor(year)2017                2.44512    0.25449  9.6079 < 2.2e-16 ***
## factor(year)2018                4.71496    0.39523 11.9296 < 2.2e-16 ***
## factor(year)2019                5.07344    0.45000 11.2744 < 2.2e-16 ***
## factor(year)2020                7.85727    0.53242 14.7576 < 2.2e-16 ***
## factor(year)2021               12.68718    0.65222 19.4523 < 2.2e-16 ***
## factor(year)2022               11.92709    0.68966 17.2942 < 2.2e-16 ***
## factor(year)2023               13.45530    0.74395 18.0862 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    330440
## Residual Sum of Squares: 132720
## R-Squared:      0.59835
## Adj. R-Squared: 0.5944
## F-statistic: 118.558 on 12 and 154 DF, p-value: < 2.22e-16
```

```r
# Adding Lag1: SPI ~ DI
ols_spi_di_L1 <- plm(
  formula = spi_comp ~ di_score + di_score_lag1 + log_gdppc + factor(income_level_recoded) + factor(yea
  index = c("country_code", "year"),
  model = "pooling",
  data = panel_data)
summary(ols_spi_di_L1, vcov = vcovHC(ols_spi_di_L1, cluster = "group", type = "HC1"))
```

```
## Pooling Model
##
## Note: Coefficient variance-covariance matrix supplied: vcovHC(ols_spi_di_L1, cluster = "group", type
##
## Call:
## plm(formula = spi_comp ~ 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 = 6-8, N = 1234
```

```
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -33.42868  -5.41171   0.82899   6.37715  26.84139
##
## Coefficients:
##                                Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)                    14.68357    9.79982  1.4984  0.134301
## di_score                       -0.27098    1.52853 -0.1773  0.859319
## di_score_lag1                   3.90609    1.48117  2.6372  0.008467 **
## log_gdppc                       2.74623    1.60865  1.7072  0.088047 .
## factor(income_level_recoded)1   0.99570    2.57320  0.3870  0.698860
## factor(income_level_recoded)2   1.63042    4.19175  0.3890  0.697375
## factor(income_level_recoded)3   3.02895    6.34042  0.4777  0.632935
## factor(year)2017                2.40269    0.25375  9.4689 < 2.2e-16 ***
## factor(year)2018                4.87629    0.40584 12.0152 < 2.2e-16 ***
## factor(year)2019                5.08210    0.44465 11.4294 < 2.2e-16 ***
## factor(year)2020                7.69359    0.52917 14.5390 < 2.2e-16 ***
## factor(year)2021               12.46231    0.63539 19.6135 < 2.2e-16 ***
## factor(year)2022               12.11244    0.69589 17.4056 < 2.2e-16 ***
## factor(year)2023               13.32427    0.72890 18.2800 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    330440
## Residual Sum of Squares: 131810
## R-Squared:      0.60111
## Adj. R-Squared: 0.59686
## F-statistic: 109.526 on 13 and 154 DF, p-value: < 2.22e-16
```

```r
# Adding Lag2: SPI ~ DI
ols_spi_di_L2 <- plm(
  formula = spi_comp ~ di_score + di_score_lag1 + di_score_lag2 + log_gdppc + #factor(income_level_reco
    factor(year),
  index = c("country_code", "year"),
  model = "pooling",
  data = panel_data)
summary(ols_spi_di_L2, vcov = vcovHC(ols_spi_di_L2, cluster = "group", type = "HC1"))
```

```
## Pooling Model
##
## Note: Coefficient variance-covariance matrix supplied: vcovHC(ols_spi_di_L2, cluster = "group", type
##
## Call:
## plm(formula = spi_comp ~ di_score + di_score_lag1 + di_score_lag2 +
##     log_gdppc + 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.
## -32.46102  -5.41476   0.76788   6.52552  27.04497
##
```

```
## Coefficients:
##                   Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)      10.05410    3.80326  2.6436 0.0083091 **
## di_score          0.13330    1.43963  0.0926 0.9262423
## di_score_lag1    -1.21009    0.81709 -1.4810 0.1388731
## di_score_lag2     4.79164    1.39728  3.4293 0.0006254 ***
## log_gdppc         3.42697    0.69684  4.9179 9.942e-07 ***
## factor(year)2017  2.17592    0.28091  7.7460 1.982e-14 ***
## factor(year)2018  4.56216    0.39891 11.4365 < 2.2e-16 ***
## factor(year)2019  5.06232    0.44417 11.3972 < 2.2e-16 ***
## factor(year)2020  7.52992    0.52559 14.3265 < 2.2e-16 ***
## factor(year)2021 12.01273    0.64827 18.5306 < 2.2e-16 ***
## factor(year)2022 11.53617    0.65295 17.6676 < 2.2e-16 ***
## factor(year)2023 13.30520    0.71505 18.6074 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    330440
## Residual Sum of Squares: 130570
## R-Squared:      0.60487
## Adj. R-Squared: 0.60132
## F-statistic: 118.317 on 11 and 154 DF, p-value: < 2.22e-16
```

3.1 POLS Summary Table

4 1.2) First Difference [Stage 1]: SPI ~ DI

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

```
## Oneway (individual) effect First-Difference Model
##
## Note: Coefficient variance-covariance matrix supplied: vcovHC(fd_spi_di, cluster = "group", type = "
##
## Call:
## plm(formula = 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.
## -10.76715  -1.97714  -0.52064   1.77908  14.85161
##
## Coefficients:
##                                Estimate Std. Error t-value Pr(>|t|)
```

```
## (Intercept)                     1.766314   0.093367 18.9180  < 2e-16 ***
## di_score                       -0.735136   0.343029 -2.1431  0.03233 *
## log_gdppc                       2.727119   0.861903  3.1641  0.00160 **
## factor(income_level_recoded)1 -0.168745   0.914976 -0.1844  0.85371
## factor(income_level_recoded)2 -1.026024   1.134772 -0.9042  0.36611
## factor(income_level_recoded)3 -2.489069   1.342032 -1.8547  0.06391 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    11236
## Residual Sum of Squares: 11086
## R-Squared:      0.013276
## Adj. R-Squared: 0.0086777
## F-statistic: 3.91929 on 5 and 154 DF, p-value: 0.0022585
```

```
# Adding Lag1: SPI ~ DI
fd_spi_di_L1 <- plm(formula = spi_comp ~ di_score + di_score_lag1 +
    log_gdppc + factor(income_level_recoded), index = c("country_code",
    "year"), data = fd_data, model = "fd")
summary(fd_spi_di_L1, vcov = vcovHC(fd_spi_di_L1, cluster = "group",
    type = "HC1"))
```

```
## Oneway (individual) effect First-Difference Model
##
## Note: Coefficient variance-covariance matrix supplied: vcovHC(fd_spi_di_L1, cluster = "group", type =
##
## Call:
## plm(formula = spi_comp ~ 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 = 6-8, N = 1234
## Observations used in estimation: 1079
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -10.97802  -1.98350  -0.54045   1.71682  14.69568
##
## Coefficients:
##                                Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)                    1.749971   0.093475 18.7213 < 2.2e-16 ***
## di_score                      -0.685783   0.356742 -1.9223  0.054827 .
## di_score_lag1                 -0.540537   0.371283 -1.4559  0.145724
## log_gdppc                      2.718265   0.863792  3.1469  0.001696 **
## factor(income_level_recoded)1 -0.140625   0.927246 -0.1517  0.879484
## factor(income_level_recoded)2 -0.975729   1.144114 -0.8528  0.393947
## factor(income_level_recoded)3 -2.477577   1.352152 -1.8323  0.067181 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    11236
## Residual Sum of Squares: 11070
## R-Squared:      0.014713
## Adj. R-Squared: 0.0091986
```

```
## F-statistic: 3.98072 on 6 and 154 DF, p-value: 0.0009876
```

```
# Adding Lag2: SPI ~ DI
fd_spi_di_L2 <- plm(formula = spi_comp ~ di_score + di_score_lag1 +
    di_score_lag2 + log_gdppc + factor(income_level_recoded),
    index = c("country_code", "year"), data = fd_data, model = "fd")
summary(fd_spi_di_L2, vcov = vcovHC(fd_spi_di_L2, cluster = "group",
    type = "HC1"))
```

```
## Oneway (individual) effect First-Difference Model
##
## Note: Coefficient variance-covariance matrix supplied: vcovHC(fd_spi_di_L2, cluster = "group", type =
##
## Call:
## plm(formula = spi_comp ~ 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 = 6-8, N = 1234
## Observations used in estimation: 1079
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -10.84779  -1.98161  -0.56005   1.67582  15.44487
##
## Coefficients:
##                                 Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)                     1.781442   0.093225 19.1090 < 2.2e-16 ***
## di_score                       -0.746543   0.361247 -2.0666  0.039015 *
## di_score_lag1                  -0.651898   0.380845 -1.7117  0.087239 .
## di_score_lag2                   1.148413   0.429147  2.6760  0.007563 **
## log_gdppc                       2.802374   0.872878  3.2105  0.001364 **
## factor(income_level_recoded)1 -0.373553   0.943423 -0.3960  0.692217
## factor(income_level_recoded)2 -1.174785   1.163739 -1.0095  0.312967
## factor(income_level_recoded)3 -2.715900   1.363316 -1.9921  0.046611 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    11236
## Residual Sum of Squares: 10997
## R-Squared:      0.021246
## Adj. R-Squared: 0.014849
## F-statistic: 3.98876 on 7 and 154 DF, p-value: 0.00048966
```

4.1 First Difference Summary Table

5 1.3) Fixed Effects [Stage 1]: SPI ~ DI

```
# Contemporaneous Effect: SPI ~ DI
fe_spi_di <- plm(formula = spi_comp ~ di_score + log_gdppc +
    factor(income_level_recoded) + factor(year), index = c("country_code",
```

```
    "year"), data = panel_data, model = "within")
summary(fe_spi_di, vcov = vcovHC(fe_spi_di, cluster = "group",
    type = "HC1"))
```

```
## Oneway (individual) effect Within Model
##
## Note: Coefficient variance-covariance matrix supplied: vcovHC(fe_spi_di, cluster = "group", type = "
##
## Call:
## plm(formula = spi_comp ~ di_score + log_gdppc + factor(income_level_recoded) +
##     factor(year), 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.
## -15.603628  -2.016964   0.026299   2.034538  15.025768
##
## Coefficients:
##                                Estimate Std. Error t-value Pr(>|t|)
## di_score                      -0.034163   0.558194 -0.0612   0.9512
## log_gdppc                      0.679409   1.713130  0.3966   0.6917
## factor(income_level_recoded)1  1.199495   2.044982  0.5866   0.5576
## factor(income_level_recoded)2 -0.239668   2.327351 -0.1030   0.9180
## factor(income_level_recoded)3  0.659352   2.489716  0.2648   0.7912
## factor(year)2017               2.565483   0.260320  9.8551   <2e-16 ***
## factor(year)2018               5.009847   0.438098 11.4354   <2e-16 ***
## factor(year)2019               5.254905   0.485108 10.8324   <2e-16 ***
## factor(year)2020               7.600822   0.532313 14.2789   <2e-16 ***
## factor(year)2021              12.334691   0.706057 17.4698   <2e-16 ***
## factor(year)2022              11.771974   0.767849 15.3311   <2e-16 ***
## factor(year)2023              12.953929   0.845892 15.3139   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    38471
## Residual Sum of Squares: 12742
## R-Squared:      0.6688
## Adj. R-Squared: 0.61727
## F-statistic: 62.2996 on 12 and 154 DF, p-value: < 2.22e-16
```

```
# Adding Lag1: SPI ~ DI
fe_spi_di_L1 <- plm(formula = spi_comp ~ di_score + di_score_lag1 +
    log_gdppc + factor(income_level_recoded) + factor(year),
    index = c("country_code", "year"), data = panel_data, model = "within")
summary(fe_spi_di_L1, vcov = vcovHC(fe_spi_di_L1, cluster = "group",
    type = "HC1"))
```

```
## Oneway (individual) effect Within Model
##
## Note: Coefficient variance-covariance matrix supplied: vcovHC(fe_spi_di_L1, cluster = "group", type =
```

```
##
## Call:
## plm(formula = spi_comp ~ 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 = 6-8, N = 1234
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -15.59202  -2.01203   0.04694   2.05282  15.00381
##
## Coefficients:
##                                Estimate Std. Error t-value Pr(>|t|)
## di_score                       -0.32493    0.57746 -0.5627   0.5738
## di_score_lag1                   0.39860    0.53726  0.7419   0.4583
## log_gdppc                       0.68130    1.71235  0.3979   0.6908
## factor(income_level_recoded)1   1.16567    2.04357  0.5704   0.5685
## factor(income_level_recoded)2 -0.27693    2.32982 -0.1189   0.9054
## factor(income_level_recoded)3   0.64238    2.48917  0.2581   0.7964
## factor(year)2017                2.56464    0.26050  9.8451   <2e-16 ***
## factor(year)2018                5.02915    0.43884 11.4601   <2e-16 ***
## factor(year)2019                5.26132    0.48590 10.8281   <2e-16 ***
## factor(year)2020                7.59767    0.53218 14.2764   <2e-16 ***
## factor(year)2021               12.33453    0.70513 17.4926   <2e-16 ***
## factor(year)2022               11.81246    0.77280 15.2853   <2e-16 ***
## factor(year)2023               12.96910    0.84679 15.3156   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    38471
## Residual Sum of Squares: 12735
## R-Squared:      0.66898
## Adj. R-Squared: 0.61712
## F-statistic: 57.9668 on 13 and 154 DF, p-value: < 2.22e-16
```

```r
# Adding Lag2: SPI ~ DI
fe_spi_di_L2 <- plm(formula = spi_comp ~ di_score + di_score_lag1 +
    di_score_lag2 + log_gdppc + factor(income_level_recoded) +
    factor(year), index = c("country_code", "year"), data = panel_data,
    model = "within")
summary(fe_spi_di_L2, vcov = vcovHC(fe_spi_di_L2, cluster = "group",
    type = "HC1"))
```

```
## Oneway (individual) effect Within Model
##
## Note: Coefficient variance-covariance matrix supplied: vcovHC(fe_spi_di_L2, cluster = "group", type =
##
## Call:
## plm(formula = spi_comp ~ 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 = 6-8, N = 1234
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -16.096338  -1.992869   0.036877   2.063381  14.903550
##
## Coefficients:
##                                 Estimate Std. Error t-value Pr(>|t|)
## di_score                        -0.24686    0.55725 -0.4430  0.65786
## di_score_lag1                   -0.42475    0.49081 -0.8654  0.38701
## di_score_lag2                    1.13841    0.64848  1.7555  0.07946 .
## log_gdppc                        0.73296    1.71605  0.4271  0.66938
## factor(income_level_recoded)1    0.96643    2.06472  0.4681  0.63983
## factor(income_level_recoded)2   -0.47203    2.37062 -0.1991  0.84221
## factor(income_level_recoded)3    0.34669    2.54279  0.1363  0.89158
## factor(year)2017                 2.52129    0.26192  9.6261  < 2e-16 ***
## factor(year)2018                 4.98625    0.43884 11.3624  < 2e-16 ***
## factor(year)2019                 5.28066    0.48628 10.8593  < 2e-16 ***
## factor(year)2020                 7.59015    0.52992 14.3232  < 2e-16 ***
## factor(year)2021                12.29857    0.70618 17.4157  < 2e-16 ***
## factor(year)2022                11.78656    0.77061 15.2950  < 2e-16 ***
## factor(year)2023                13.06564    0.83810 15.5896  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    38471
## Residual Sum of Squares: 12683
## R-Squared:      0.67033
## Adj. R-Squared: 0.61832
## F-statistic: 53.0318 on 14 and 154 DF, p-value: < 2.22e-16
```

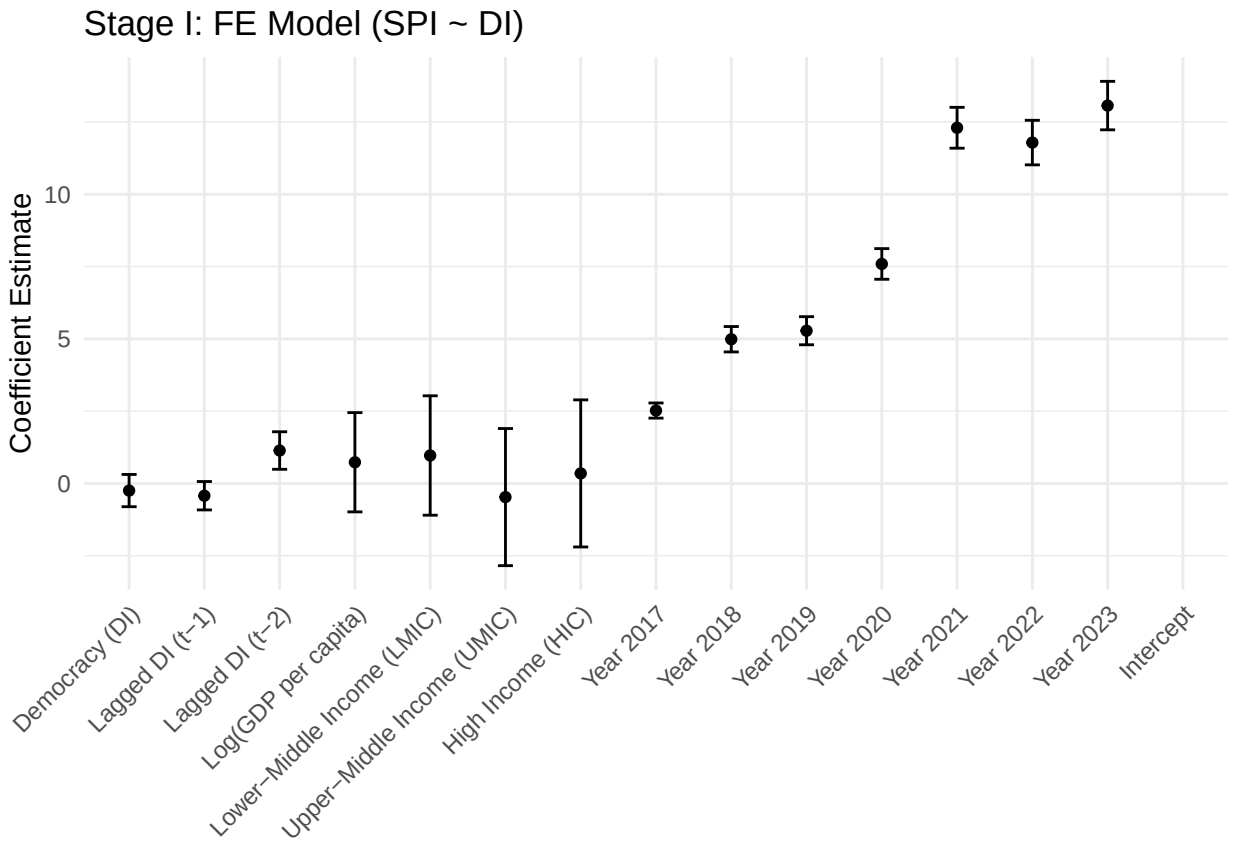
5.1 Fixed Effects Summary Table

5.1.1 stargazer table for only lag2 models

5.2 Fixed Effects Error Bar Visualization

```r
# Extract coefficients and robust standard errors from the
# FE model
fe_spi_di_L2_results <- summary(fe_spi_di_L2, vcov = vcovHC(fe_spi_di_L2,
    cluster = "group", type = "HC1"))
# Create a data frame for visualization
coef_df <- data.frame(term = rownames(fe_spi_di_L2_results$coefficients),
    estimate = fe_spi_di_L2_results$coefficients[, "Estimate"],
    std.error = fe_spi_di_L2_results$coefficients[, "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 I: FE Model (SPI ~ DI)",
    x = NULL, y = "Coefficient Estimate") + theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
    # re-labeling x variable lables
```

```
scale_x_discrete(labels = c(di_score = "Democracy (DI)", di_score_lag1 = "Lagged DI (t-1)",
    di_score_lag2 = "Lagged DI (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", "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
```



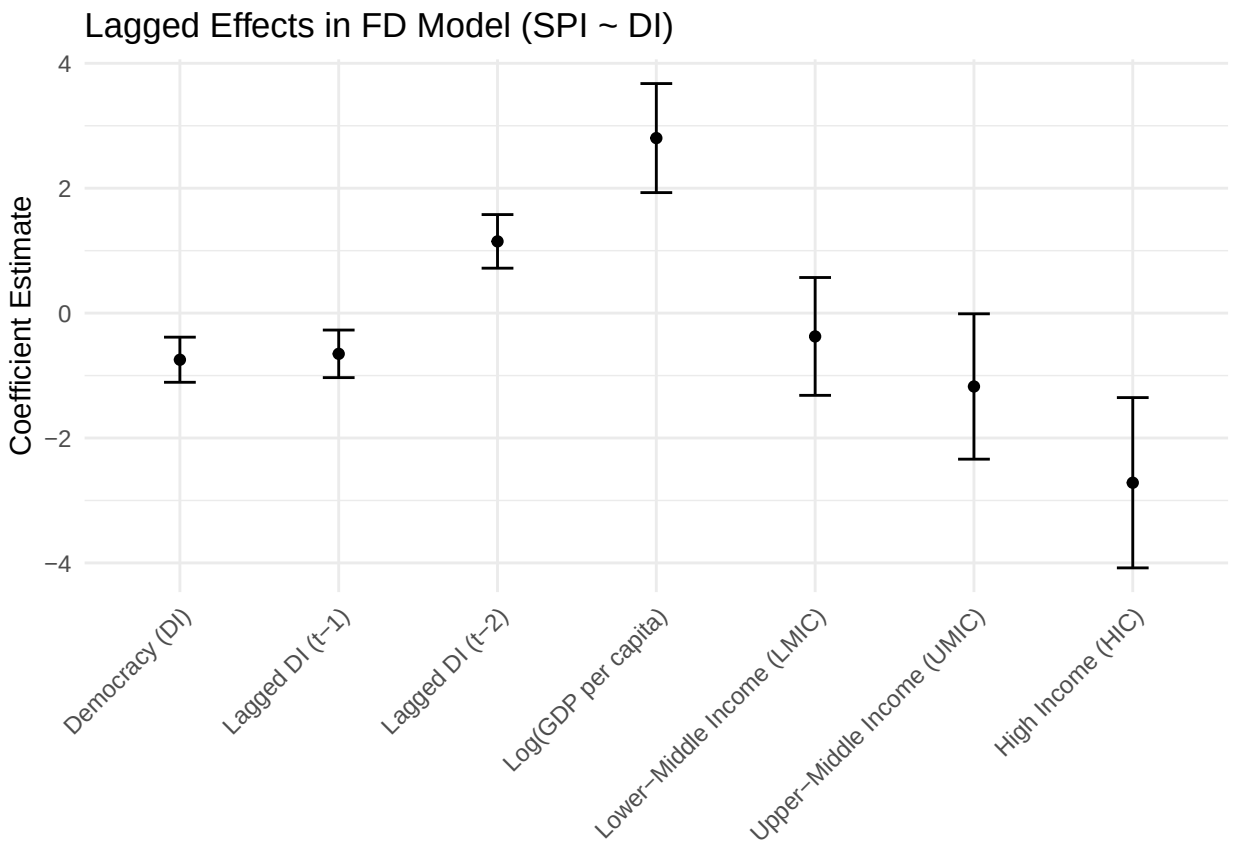

```
# Save the plot
# ggsave('component_2/figures/stage1/error_bar_fe_spi_di_L2.png',
# ebar_fe, width = 10, height = 6)
```

5.3 First Difference Error Bar Visualization

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

```r
fd_spi_di_L2_results <- summary(fd_spi_di_L2, vcov = vcovHC(fd_spi_di_L2,
    cluster = "group", type = "HC1"))
# Create a data frame for visualization
coef_fd_df <- data.frame(term = rownames(fd_spi_di_L2_results$coefficients),
    estimate = fd_spi_di_L2_results$coefficients[, "Estimate"],
    std.error = fd_spi_di_L2_results$coefficients[, "Std. Error"])
# Create a ggplot error bar chart
ebar_fd <- ggplot(coef_fd_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 FD Model (SPI ~ DI)",
    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)",
        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)"),
        limits = c("di_score", "di_score_lag1", "di_score_lag2",
            "log_gdppc", "factor(income_level_recoded)1", "factor(income_level_recoded)2",
            "factor(income_level_recoded)3"))

ebar_fd
```




```r
# Save the plot
# ggsave('component_2/figures/stage1/error_bar_fd_spi_di_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_spi_di_L2)  # [significant]
```

```
##
##  Wooldridge's test for serial correlation in FE panels
##
## data:  fe_spi_di_L2
## F = 500.43, df1 = 1, df2 = 1077, 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.

This is corrected by using robust standard errors clustered by country, which accounts for the potential autocorrelation in the residuals.

5.5 Check for Heteroskedasticity

```
# Apply Breusch-Pagan test for heteroskedasticity
bptest(fe_spi_di_L2, studentize = TRUE)  # Heteroskedasticity [significant]
```

```
##
##  studentized Breusch-Pagan test
##
## data:  fe_spi_di_L2
## BP = 139.39, df = 14, p-value < 2.2e-16
```

```
bptest(fd_spi_di_L2, studentize = TRUE)  # Heteroskedasticity [significant]
```

```
##
##  studentized Breusch-Pagan test
##
## data:  fd_spi_di_L2
## BP = 133.37, 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.