Component 2, Stage 1

Sevastian Sanchez

2025-10-13

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

1 Set up

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

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

# load data
source("Comp2_panel_wrangling.R")
panel_data1 <- panel_data %>%
    dplyr::select(country_name, country_code, year, spi_comp,
        di_score, log_gdppc, income_level_recoded, di_score_reverse) %>%
    dplyr::arrange(country_code, year)

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

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

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

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

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

sensitivity analysis [REVISE -USING PLM::LAGs] 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

2.1 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
```

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),
  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(year),
##     data = panel_data, model = "pooling")
##
## Unbalanced Panel: n = 155, T = 6-8, N = 1234
##
## Residuals:
```

```
##       Min.   1st Qu.   Median   3rd Qu.      Max.
## -32.83487  -5.58135   0.82944   6.50077  27.89241
##
## Coefficients:
##                   Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)       10.90623    3.89985  2.7966  0.005246 **
## di_score           3.57390    0.55766  6.4087 2.089e-10 ***
## log_gdppc          3.42650    0.71356  4.8020 1.765e-06 ***
## factor(year)2017   2.41527    0.25917  9.3192 < 2.2e-16 ***
## factor(year)2018   4.66006    0.39591 11.7705 < 2.2e-16 ***
## factor(year)2019   5.05045    0.45343 11.1383 < 2.2e-16 ***
## factor(year)2020   7.86974    0.53008 14.8463 < 2.2e-16 ***
## factor(year)2021  12.61948    0.65597 19.2378 < 2.2e-16 ***
## factor(year)2022  11.84161    0.68579 17.2671 < 2.2e-16 ***
## factor(year)2023  13.37865    0.74154 18.0417 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:     330440
## Residual Sum of Squares: 132890
## R-Squared:        0.59785
## Adj. R-Squared: 0.59489
## F-statistic: 149.872 on 9 and 154 DF, p-value: < 2.22e-16
```

```r
# Adding Lag1: SPI ~ DI
ols_spi_di_L1 <- plm(
  formula = spi_comp ~ di_score + plm::lag(di_score, 1)
  + log_gdppc
  #+ factor(income_level_recoded)
  + factor(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 + plm::lag(di_score, 1) + log_gdppc +
##     factor(year), data = panel_data, model = "pooling")
##
## Unbalanced Panel: n = 155, T = 6-7, N = 1082
##
## Residuals:
##       Min.   1st Qu.   Median   3rd Qu.      Max.
## -32.38842  -5.21201   0.91248   6.28462  25.86064
##
## Coefficients:
##                       Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)           12.49441    3.94091  3.1704  0.001565 **
## di_score               0.27964    1.56482  0.1787  0.858201
## plm::lag(di_score, 1)  3.21368    1.52475  2.1077  0.035291 *
## log_gdppc              3.55933    0.71171  5.0011 6.655e-07 ***
```

3

```
## factor(year)2018        2.40687      0.31168  7.7223 2.611e-14 ***
## factor(year)2019        2.66471      0.37015  7.1991 1.139e-12 ***
## factor(year)2020        5.33644      0.46283 11.5300 < 2.2e-16 ***
## factor(year)2021       10.01367      0.55913 17.9094 < 2.2e-16 ***
## factor(year)2022        9.56654      0.60823 15.7285 < 2.2e-16 ***
## factor(year)2023       10.83316      0.64972 16.6735 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:     279550
## Residual Sum of Squares: 112770
## R-Squared:        0.59661
## Adj. R-Squared: 0.59322
## F-statistic: 128.403 on 9 and 154 DF, p-value: < 2.22e-16
```

```r
# Adding Lag2: SPI ~ DI
ols_spi_di_L2 <- plm(
  formula = spi_comp ~ 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_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 + plm::lag(di_score, 1) + plm::lag(di_score,
##      2) + log_gdppc + factor(year), data = panel_data, model = "pooling")
##
## Unbalanced Panel: n = 155, T = 5-6, N = 927
##
## Residuals:
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -32.3217  -5.2488   1.0721   6.4803  25.1165
##
## Coefficients:
##                       Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)           13.46484    3.90965  3.4440 0.0005991 ***
## di_score               0.68247    1.48490  0.4596 0.6459095
## plm::lag(di_score, 1) -1.51143    0.86389 -1.7496 0.0805303 .
## plm::lag(di_score, 2)  4.25058    1.61197  2.6369 0.0085088 **
## log_gdppc              3.75030    0.71236  5.2646 1.75e-07 ***
## factor(year)2019       0.49798    0.23192  2.1472 0.0320377 *
## factor(year)2020       3.02739    0.34616  8.7457 < 2.2e-16 ***
## factor(year)2021       7.48329    0.48479 15.4360 < 2.2e-16 ***
## factor(year)2022       6.90966    0.50016 13.8148 < 2.2e-16 ***
## factor(year)2023       8.65912    0.60242 14.3739 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

4

```
## Total Sum of Squares:     233500
## Residual Sum of Squares: 93657
## R-Squared:       0.5989
## Adj. R-Squared: 0.59496
## F-statistic: 116.61 on 9 and 154 DF, p-value: < 2.22e-16
```

3.1 POLS Scatterplots

General relationship between SPI and DI

```
# Contemporaneous Relationship: SPI ~ DI
spi_di_s1_scatter <- ggplot(panel_data1, aes(x = di_score, y = spi_comp)) +
    geom_point(color = "steelblue4", size = 1, alpha = 0.65) +
    geom_smooth(method = "lm", se = TRUE, color = "darkblue",
        size = 1) + labs(title = "Effect of Democracy Levels on Statistical Capacity",
    x = "Democracy Index (0-10 Scale)", y = "SPI Composite (0-100 Scale)") +
    theme_minimal() + theme(plot.title = element_text(size = 14)) +
    ylim(20, 100)

# Save to specific folder
# ggsave('figures/stage_1_n_2_scatterplots/spi_di_s1_scatterplot.png',
# spi_di_s1_scatter, width = 6, height = 8)

# Contemporaneous Relationship: SPI ~ DI_Reversed
spi_di_rev_scatter <- ggplot(panel_data1, aes(x = di_score_reverse,
    y = spi_comp)) + geom_point(color = "steelblue4", size = 1,
    alpha = 0.65) + geom_smooth(method = "lm", se = TRUE, color = "darkblue",
    size = 1) + labs(title = "Effect of Autocracy Levels on Statistical Capacity (reversed DI)",
    x = "Democracy Index (0-10 Scale) [Higher = More Autocratic]",
    y = "SPI Composite (0-100 Scale)") + theme_minimal() + theme(plot.title = element_text(size = 14))
    # adjust y axis to start from y = 10
ylim(20, 100)

# Save to specific folder
# ggsave('figures/stage_1_n_2_scatterplots/spi_di_rev_scatter.png',
# spi_di_rev_scatter, width = 6, height = 8)

# side by side comparison using patchwork
library(patchwork)
spi_di_s1_scatter + spi_di_rev_scatter + plot_layout(ncol = 2)
```
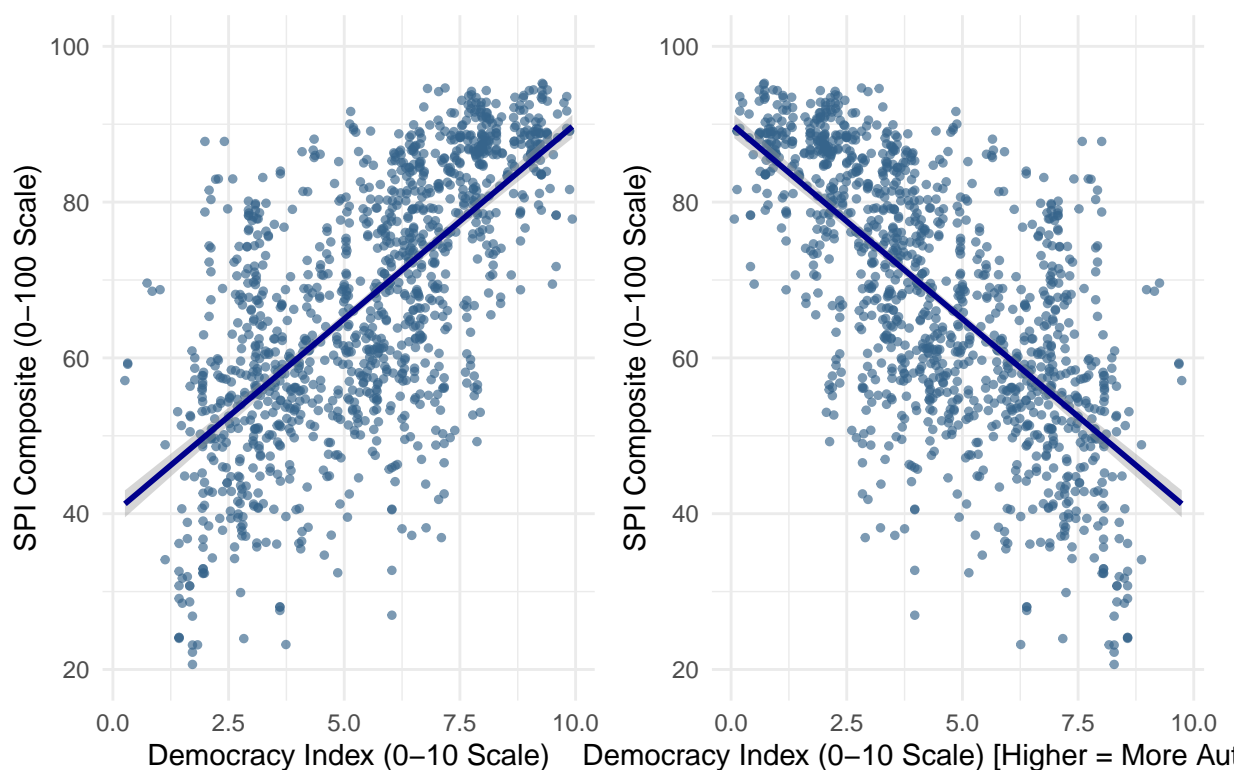
Effect of Democracy Levels on Statistical Capacity / Effect of Democracy Levels on St...

```
# ggsave('figures/stage_1_n_2_scatterplots/spi_di_s1_scatterplots.png',
# width = 12, height = 8)
```

3.2 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),
  #+ factor(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 = "H
##
## Call:
## plm(formula = 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.
## -10.76962  -1.97302  -0.53025   1.73174  14.87242
##
## Coefficients:
##               Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)  1.763329   0.093767 18.8055 < 2.2e-16 ***
## di_score    -0.703689   0.333219 -2.1118  0.034934 *
## log_gdppc    2.466664   0.841780  2.9303  0.003458 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:     11236
## Residual Sum of Squares: 11129
## R-Squared:       0.009438
## Adj. R-Squared: 0.0075968
## F-statistic: 5.91662 on 2 and 154 DF, p-value: 0.0033449
```

```r
# Adding Lag1: SPI ~ DI
fd_spi_di_L1 <- plm(
  formula = spi_comp ~ di_score + plm::lag(di_score, 1)
  + log_gdppc,
  #+ factor(income_level_recoded),
  #+ factor(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 + 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.
## -10.89959  -1.93429  -0.53814   1.66731  14.83495
##
## Coefficients:
##                        Estimate Std. Error t-value Pr(>|t|)
## (Intercept)            1.630254   0.098787 16.5027  < 2e-16 ***
## di_score              -0.817510   0.354976 -2.3030  0.02150 *
## plm::lag(di_score, 1) -0.620626   0.379412 -1.6358  0.10223
```

```
## log_gdppc                    2.352569   1.000289   2.3519   0.01889 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:      10003
## Residual Sum of Squares: 9881.1
## R-Squared:       0.012168
## Adj. R-Squared: 0.0089467
## F-statistic: 5.86215 on 3 and 154 DF, p-value: 0.00081221
```

```
# Adding Lag2: SPI ~ DI
fd_spi_di_L2 <- plm(
  formula = spi_comp ~ di_score + plm::lag(di_score, 1) + plm::lag(di_score, 2)
  + log_gdppc,
  #+ factor(income_level_recoded),
  #+ factor(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 + 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.
## -10.64368  -1.84234  -0.44069   1.57447  15.82469
##
## Coefficients:
##                        Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)            1.515191   0.095169 15.9211 < 2.2e-16 ***
## di_score              -1.310760   0.371455 -3.5287 0.0004425 ***
## plm::lag(di_score, 1) -0.887450   0.443631 -2.0004 0.0458079 *
## plm::lag(di_score, 2)  1.338215   0.451396  2.9646 0.0031249 **
## log_gdppc              2.927504   0.922686  3.1728 0.0015701 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:      7724.4
## Residual Sum of Squares: 7458.6
## R-Squared:       0.03441
## Adj. R-Squared: 0.029354
## F-statistic: 8.5748 on 4 and 154 DF, p-value: 2.8353e-06
```

```
# CI for all variables (L2 model)
coef <- coef(fd_spi_di_L2)
se <- sqrt(diag(vcovHC(fd_spi_di_L2, cluster = "group", type = "HC1")))  # Get standard errors
crit_value <- qt(0.975, df = nobs(fd_spi_di_L2)-length(coef(fd_spi_di_L2)))
ci_low <- coef - crit_value * se
ci_high <- coef + crit_value * se
cbind(ci_low, ci_high)
```

```
##                              ci_low     ci_high
## (Intercept)              1.3283672   1.7020144
## di_score                -2.0399522  -0.5815669
## plm::lag(di_score, 1)   -1.7583310  -0.0165698
## plm::lag(di_score, 2)    0.4520909   2.2243398
## log_gdppc                1.1162039   4.7388037
```

```
# joint significance
# names(coef(fd_spi_di_L2)) # ensure correct names
linearHypothesis(
  fd_spi_di_L2,
  c("di_score = 0", "plm::lag(di_score, 1) = 0", "plm::lag(di_score, 2) = 0"),
  vcov. = vcovHC(fd_spi_di_L2, cluster = "group", type = "HC1")
)
```

```
##
## Linear hypothesis test:
## di_score = 0
## plm::lag(di_score,0
## plm::lag(di_score, 2) = 0
##
## Model 1: restricted model
## Model 2: spi_comp ~ di_score + plm::lag(di_score, 1) + plm::lag(di_score,
##      2) + log_gdppc
##
## Note: Coefficient covariance matrix supplied.
##
##   Res.Df Df  Chisq Pr(>Chisq)
## 1    767
## 2    764  3 25.898  1.002e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# cumulative effects
linearHypothesis(
  fd_spi_di_L2,
  c("di_score + plm::lag(di_score, 1) + plm::lag(di_score, 2) = 0"),
  vcov. = vcovHC(fd_spi_di_L2, cluster = "group", type = "HC1")
)
```

```
##
## Linear hypothesis test:
## di_score  + plm::lag(di_score,1 plm::lag(di_score, 2) = 0
##
```

9

```
## Model 1: restricted model
## Model 2: spi_comp ~ di_score + plm::lag(di_score, 1) + plm::lag(di_score,
##     2) + log_gdppc
##
## Note: Coefficient covariance matrix supplied.
##
##   Res.Df Df  Chisq Pr(>Chisq)
## 1    765
## 2    764  1 2.5181     0.1125
```

Result: The joint significance tests for FD show that the DI variables are jointly significant in explaining changes in SPI overall scores (Chisq = 25.898, p = 1.002e-05). This suggests that shifts in democracy levels may have a meaningful immediate impact on statistical capacity (SPI) performance when considering both current and lagged effects.

In terms of *cumulative impact*, based on the FD model and panel data, there is no detectable total impact of changes in di_score (at current, lag 1, or lag 2) on SPI —over these periods (stat = 2.5181 p = 0.1125)

4.1   First Difference Summary Table

4.2   First Difference Error Bar Visualization

```r
# Extract coefficients and robust standard errors from the
# FD model
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 = "Static & Lagged First Difference (SPI~DI
    x = NULL, y = "Coefficient Estimate") + theme_minimal() +
    coord_flip() + geom_hline(yintercept = 0, linetype = "dashed",
    color = "steelblue") + 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)",
    log_gdppc = "Log(GDP per capita)", `(Intercept)` = "Intercept"),
    limits = c("(Intercept)", "log_gdppc", "plm::lag(di_score, 2)",
        "plm::lag(di_score, 1)", "di_score"))
ebar_fd
```

Static & Lagged First Difference (SPI~DI)
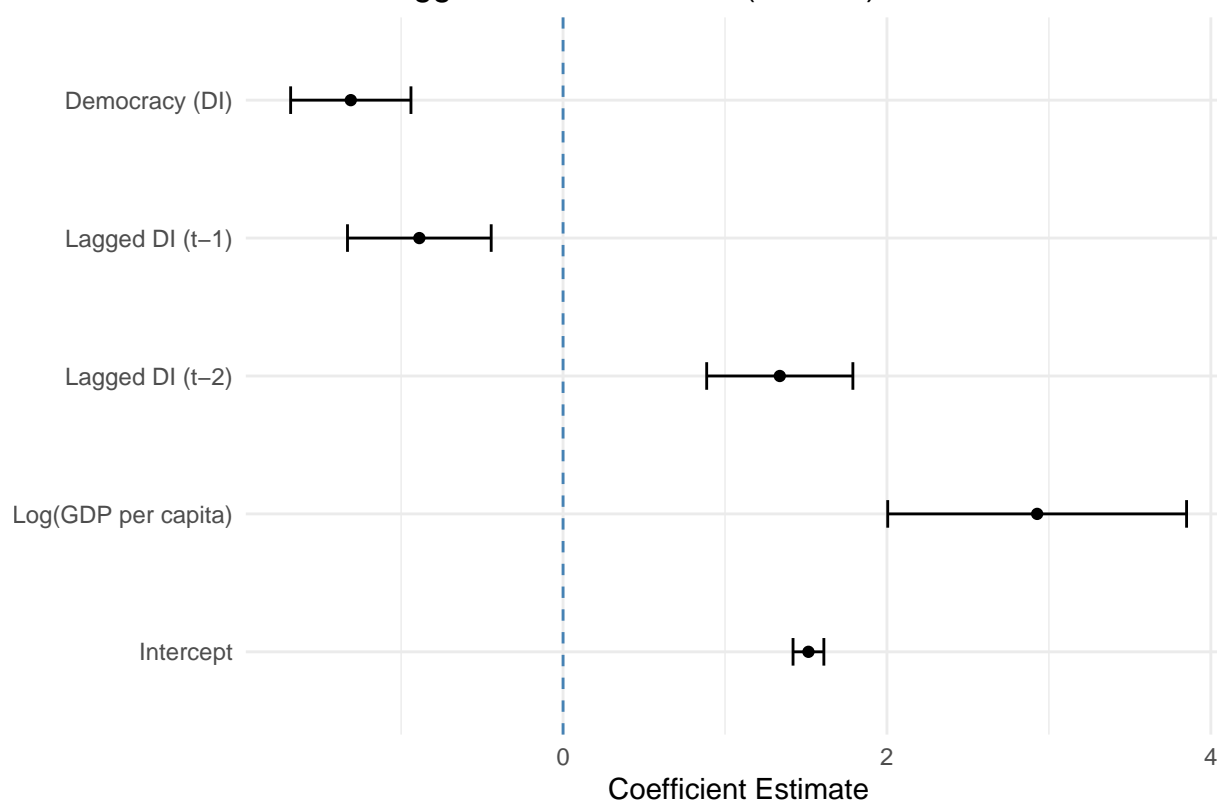

```
# Save the plot
# ggsave('component_2/figures/stage1/error_bar_fd_spi_di_L2.png',
# ebar_fd, width = 10, height = 6)
```

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),
  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 = "H
##
## Call:
## plm(formula = 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.
## -15.599271  -2.012044   0.026745   2.034042  15.044393
## 
## Coefficients:
##                    Estimate Std. Error t-value Pr(>|t|)
## di_score          -0.098521   0.550666 -0.1789   0.8580
## log_gdppc          0.614503   1.732569  0.3547   0.7229
## factor(year)2017   2.537886   0.260767  9.7324   <2e-16 ***
## factor(year)2018   4.977623   0.447103 11.1331   <2e-16 ***
## factor(year)2019   5.257074   0.498000 10.5564   <2e-16 ***
## factor(year)2020   7.600929   0.536622 14.1644   <2e-16 ***
## factor(year)2021  12.349788   0.721917 17.1069   <2e-16 ***
## factor(year)2022  11.804139   0.790076 14.9405   <2e-16 ***
## factor(year)2023  12.952828   0.867039 14.9392   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Total Sum of Squares:     38471
## Residual Sum of Squares: 12811
## R-Squared:      0.66699
## Adj. R-Squared: 0.61626
## F-statistic: 77.6701 on 9 and 154 DF, p-value: < 2.22e-16
```

```r
# Adding Lag1: SPI ~ DI
fe_spi_di_L1 <- plm(
  formula = spi_comp ~ di_score + plm::lag(di_score, 1)
  + log_gdppc
  #+ factor(income_level_recoded)
  + factor(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 + plm::lag(di_score, 1) + log_gdppc +
##     factor(year), data = panel_data, model = "within")
## 
## Unbalanced Panel: n = 155, T = 6-7, N = 1082
## 
## Residuals:
##        Min.    1st Qu.     Median    3rd Qu.        Max.
## -14.330754  -1.686187   0.042262   1.814537  13.057740
## 
## Coefficients:
##                         Estimate Std. Error t-value  Pr(>|t|)
## di_score                -0.30937    0.52763 -0.5863    0.5578
```

```
## plm::lag(di_score, 1)  0.15005      0.51773  0.2898     0.7720
## log_gdppc               0.38938      1.60487  0.2426     0.8083
## factor(year)2018        2.46051      0.31754  7.7487 2.455e-14 ***
## factor(year)2019        2.73176      0.38172  7.1565 1.693e-12 ***
## factor(year)2020        5.04823      0.44937 11.2340 < 2.2e-16 ***
## factor(year)2021        9.81600      0.57936 16.9428 < 2.2e-16 ***
## factor(year)2022        9.30148      0.63101 14.7405 < 2.2e-16 ***
## factor(year)2023       10.45690      0.68905 15.1759 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    25920
## Residual Sum of Squares: 9413.8
## R-Squared:      0.63681
## Adj. R-Squared: 0.57232
## F-statistic: 64.242 on 9 and 154 DF, p-value: < 2.22e-16
```

```r
# Adding Lag2: SPI ~ DI
fe_spi_di_L2 <- plm(
  formula = spi_comp ~ di_score + plm::lag(di_score, 1) + plm::lag(di_score, 2)
  + log_gdppc
  #+ factor(income_level_recoded)
  + factor(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 + plm::lag(di_score, 1) + plm::lag(di_score,
##     2) + log_gdppc + factor(year), data = panel_data, model = "within")
##
## Unbalanced Panel: n = 155, T = 5-6, N = 927
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -12.253528  -1.371080  -0.066626   1.562429  10.444775
##
## Coefficients:
##                       Estimate Std. Error t-value  Pr(>|t|)
## di_score              -0.44904    0.52526 -0.8549    0.3929
## plm::lag(di_score, 1) -0.33377    0.36824 -0.9064    0.3650
## plm::lag(di_score, 2)  0.71220    0.55023  1.2944    0.1959
## log_gdppc             -0.11182    1.14894 -0.0973    0.9225
## factor(year)2019       0.30110    0.21130  1.4250    0.1546
## factor(year)2020       2.54965    0.33585  7.5916 9.234e-14 ***
## factor(year)2021       7.34821    0.42952 17.1081 < 2.2e-16 ***
## factor(year)2022       6.87782    0.45847 15.0017 < 2.2e-16 ***
## factor(year)2023       8.12648    0.50809 15.9940 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Total Sum of Squares:     16634
## Residual Sum of Squares: 6022.2
## R-Squared:      0.63797
## Adj. R-Squared: 0.56063
## F-statistic: 57.9356 on 9 and 154 DF, p-value: < 2.22e-16
```

```r
# CI for all variables (L2 model)
coef <- coef(fe_spi_di_L2)
se <- sqrt(diag(vcovHC(fe_spi_di_L2, cluster = "group", type = "HC1")))  # Get standard errors
crit_value <- qt(0.975, df = nobs(fe_spi_di_L2)-length(coef(fe_spi_di_L2)))
ci_low <- coef - crit_value * se
ci_high <- coef + crit_value * se
cbind(ci_low, ci_high)
```

```
##                          ci_low    ci_high
## di_score             -1.4798888 0.5818183
## plm::lag(di_score, 1) -1.0564566 0.3889078
## plm::lag(di_score, 2) -0.3676464 1.7920541
## log_gdppc            -2.3666720 2.1430335
## factor(year)2019     -0.1135954 0.7157957
## factor(year)2020      1.8905255 3.2087737
## factor(year)2021      6.5052612 8.1911589
## factor(year)2022      5.9780503 7.7775859
## factor(year)2023      7.1293194 9.1236413
```

```r
# joint significance: is there an effect between these three variables?
linearHypothesis(
  fe_spi_di_L2,
  c("di_score = 0", "plm::lag(di_score, 1) = 0", "plm::lag(di_score, 2) = 0"),
  vcov. = vcovHC(fe_spi_di_L2, cluster = "group", type = "HC1")
)
```

```
##
## Linear hypothesis test:
## di_score = 0
## plm::lag(di_score,0
## plm::lag(di_score, 2) = 0
##
## Model 1: restricted model
## Model 2: spi_comp ~ di_score + plm::lag(di_score, 1) + plm::lag(di_score,
##     2) + log_gdppc + factor(year)
##
## Note: Coefficient covariance matrix supplied.
##
##   Res.Df Df  Chisq Pr(>Chisq)
## 1    766
## 2    763  3 3.5345     0.3163
```

```r
# ensure correct names
# names(coef(fe_spi_di_L2))
```

```r
# cumulative effects
linearHypothesis(
  fe_spi_di_L2,
  c("di_score + plm::lag(di_score, 1) + plm::lag(di_score, 2) = 0"),
  vcov. = vcovHC(fe_spi_di_L2, cluster = "group", type = "HC1")
)
```

```
##
## Linear hypothesis test:
## di_score  + plm::lag(di_score,1 plm::lag(di_score, 2) = 0
##
## Model 1: restricted model
## Model 2: spi_comp ~ di_score + plm::lag(di_score, 1) + plm::lag(di_score,
##     2) + log_gdppc + factor(year)
##
## Note: Coefficient covariance matrix supplied.
##
##   Res.Df Df  Chisq Pr(>Chisq)
## 1    764
## 2    763  1 0.0107     0.9175
```

Results: The joint significance tests for FE show no evidence that DI variables are jointly significant in explaining changes in SPI scores (Chisq = 3.5345, p = 0.3163). This suggests that DI may not have a meaningful incremental impact on SDG outcomes when considering both current and lagged effects.

In terms of *cumulative impact*, based on the FE model and panel data, there is no detectable total impact of changes in di_score (at current, lag 1, or lag 2) on SPI —over these periods. (stat = 0.0107, p = 0.9175)

5.1   Fixed Effects Summary Table

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 = "HC
# 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: Static & Lagged Two-Way Fixed Effects (SPI ~ DI)",
       x = NULL,
       y = "Coefficient Estimate") +
  theme_minimal() +
  coord_flip() +  # Flip coordinates for better readability
  geom_hline(yintercept = 0, linetype = "dashed", color = "steelblue") +
  # re-labeling x variable lables
```

```
    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)",
                                "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(
                     "(Intercept)",
                     "factor(year)2023",
                     "factor(year)2022",
                     "factor(year)2021",
                     "factor(year)2020",
                     "factor(year)2019",
                     "factor(year)2018",
                     "factor(year)2017",
                     "log_gdppc",
                     "plm::lag(di_score, 2)",
                     "plm::lag(di_score, 1)",
                     "di_score"
                     ))
ebar_fe
```

Stage I: Static & Lagged Two−Way Fixed Effects (SPI ~ DI)
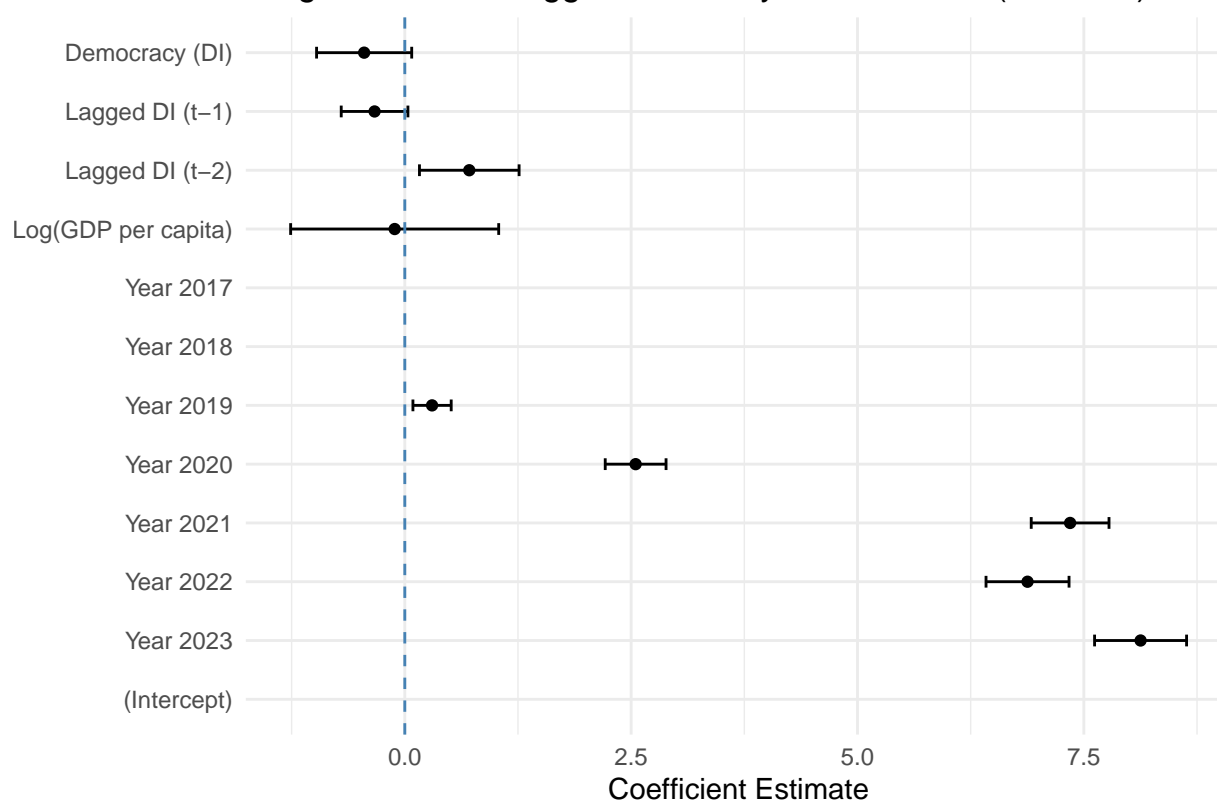

```r
# Save the plot
#ggsave("component_2/figures/stage1/error_bar_fe_spi_di_L2.png", ebar_fe, width = 10, height = 6)
```

5.2.1 stargazer table for only lag2 models

5.3 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!!!
pbgtest(fe_spi_di_L2)  # Panel
```

```
##
##  Breusch-Godfrey/Wooldridge test for serial correlation in panel models
##
## data:  spi_comp ~ di_score + plm::lag(di_score, 1) + plm::lag(di_score,  ...
## chisq = 199.21, df = 5, p-value < 2.2e-16
## alternative hypothesis: serial correlation in idiosyncratic errors
```

```r
pwartest(fe_spi_di_L2)  # FE [significant]
```

```
##
##  Wooldridge's test for serial correlation in FE panels
##
## data:  fe_spi_di_L2
## F = 322.27, df1 = 1, df2 = 770, p-value < 2.2e-16
## alternative hypothesis: serial correlation
```

```
pwfdtest(fd_spi_di_L2)  # FD [significant]
```

```
##
##  Wooldridge's first-difference test for serial correlation in panels
##
## data:  fd_spi_di_L2
## F = 5.2524, df1 = 1, df2 = 612, p-value = 0.02225
## alternative hypothesis: serial correlation in differenced errors
```

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.

5.4   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 = 64.177, df = 9, p-value = 2.085e-10
```

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

```
##
##  studentized Breusch-Pagan test
##
## data:  fd_spi_di_L2
## BP = 62.537, df = 4, p-value = 8.495e-13
```

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.

5.5   Residual Diagnostics

5.5.1   PLOS Residuals for mediator model: SPI ~ DI + Controls [STAGE 1]

```r
# Extract the data actually used in the model
ols_spi_di <- lm(spi_comp ~ di_score + log_gdppc + factor(year),
    data = panel_data)  # regular linear model
model_data1 <- model.frame(ols_spi_di)

# switch to all stage-one plots
png("component_2/figures/stage1/all_s1_residual_plots.png", width = 8,
    height = 6.5, units = "in", res = 300)
par(mfrow = c(3, 3))

# Residuals vs Fitted Values plot for ols_spi_di
plot(ols_spi_di, which = 1, main = "Residuals vs Fitted Model",
    caption = "ols_spi_di [Stage 1]", pch = 1, cex = 0.35, col = "#595959")
abline(h = 0, col = "black", lty = 2, lwd = 1)
fitted_vals <- fitted(ols_spi_di)
lines(lowess(fitted_vals, resid(ols_spi_di)), col = "red", lwd = 1.5)

# Residuals vs di_score plot
plot(model_data1$di_score, resid(ols_spi_di), xlab = "di_score",
    ylab = "", yaxt = "n", main = "Residuals vs di_score", pch = 1,
    cex = 0.35, col = "darkred")
abline(h = 0, col = "black", lty = 2, lwd = 1)
lines(lowess(model_data1$di_score, resid(ols_spi_di)), col = "red",
    lwd = 1.5)

# Residuals vs log_gdppc plot
plot(model_data1$log_gdppc, resid(ols_spi_di), xlab = "log_gdppc",
    ylab = "", yaxt = "n", main = "Residuals vs log_gdppc", pch = 1,
    cex = 0.35, col = "darkgreen")
abline(h = 0, col = "black", lty = 2, lwd = 1)
lines(lowess(model_data1$log_gdppc, resid(ols_spi_di)), col = "red",
    lwd = 1.5)

# dev.off() # to save

# par(mfrow = c(1, 1))
```

### 5.5.2 First Differences Residuals for mediator model: SPI ~ DI + Controls [STAGE 1]

```r
# Extract the data actually used in the model
fd_spi_di <- lm(spi_diff ~ di_diff + log_gdppc_diff, data = fd_data)
model_data_fd1 <- model.frame(fd_spi_di)

# saving as png
# png('figures/residual_plots/fd1_residual_plots.png',
# width = 8, height = 3, units = 'in', res = 300) par(mfrow
# = c(1, 3))

# Residuals vs Fitted Values plot for fd_spi_di
plot(fd_spi_di, which = 1, main = "FD Residuals vs Fitted Model",
    caption = "fd_spi_di [Stage 1]", pch = 1, cex = 0.35, col = "#595959")
```

```
abline(h = 0, col = "black", lty = 2, lwd = 1)
fitted_vals_fd <- fitted(fd_spi_di)
lines(lowess(fitted_vals_fd, resid(fd_spi_di)), col = "red",
    lwd = 1.5)
```

FD Residuals vs Fitted Model

fd_spi_di [Stage 1]

Fitted values
lm(spi_diff ~ di_diff + log_gdppc_diff)
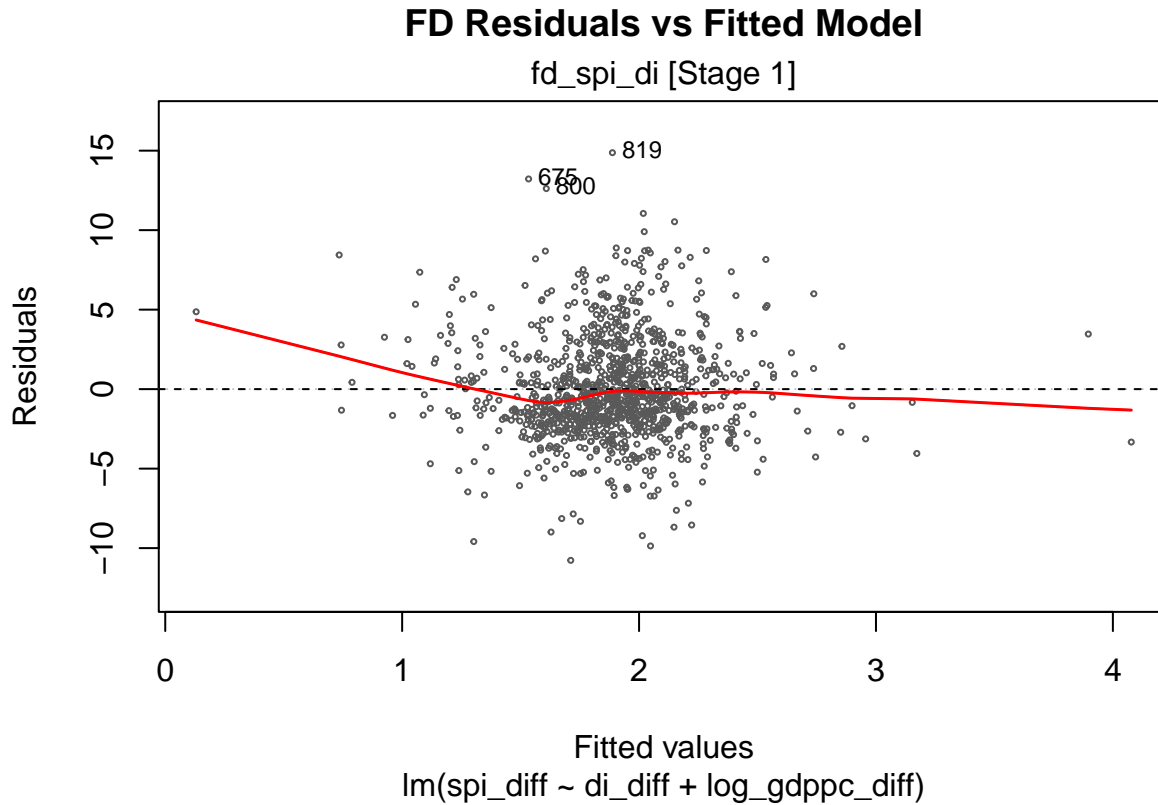
```
# Residuals vs di_diff plot
plot(model_data_fd1$di_diff, resid(fd_spi_di), xlab = "di_diff",
    ylab = "", yaxt = "n", main = "FD Residuals vs di_diff",
    pch = 1, cex = 0.35, col = "darkred")
abline(h = 0, col = "black", lty = 2, lwd = 1)
lines(lowess(model_data_fd1$di_diff, resid(fd_spi_di)), col = "red",
    lwd = 1.5)
```

FD Residuals vs di_diff

di_diff
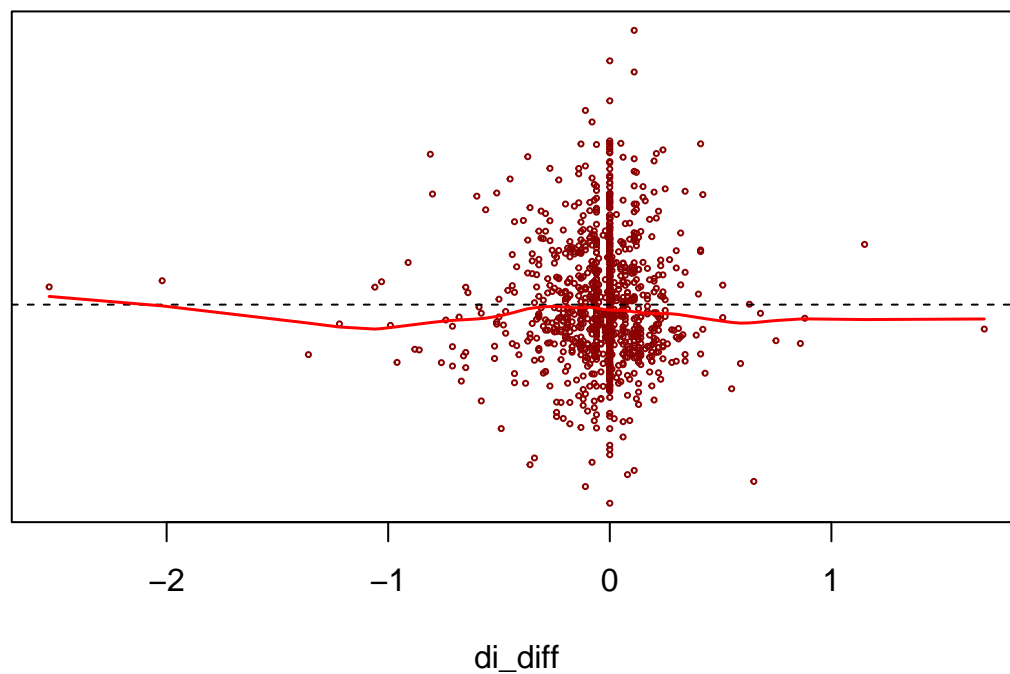
```r
# Residuals vs log_gdppc_diff plot
plot(model_data_fd1$log_gdppc_diff, resid(fd_spi_di), xlab = "log_gdppc_diff",
    ylab = "", yaxt = "n", main = "FD Residuals vs log_gdppc_diff",
    pch = 1, cex = 0.35, col = "darkgreen")
abline(h = 0, col = "black", lty = 2, lwd = 1)
lines(lowess(model_data_fd1$log_gdppc_diff, resid(fd_spi_di)),
    col = "red", lwd = 1.5)
```

FD Residuals vs log_gdppc_diff
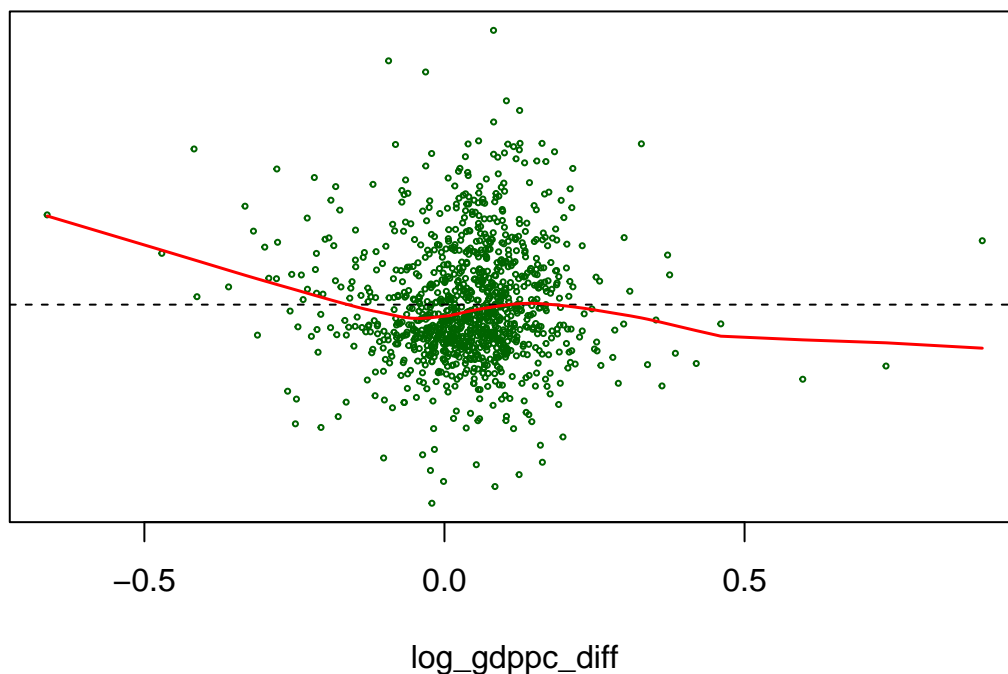


```
# dev.off() # to save par(mfrow = c(1, 1))
```

5.5.3 FE Residuals for mediator model: SPI ~ DI + Controls [STAGE 1]

```
# Extract the data actually used in the model
fe_spi_di <- lm(spi_comp ~ di_score + log_gdppc + factor(year) +
    factor(country_code), data = panel_data)
model_data_fe1 <- model.frame(fe_spi_di)

# saving as png
# png('figures/residual_plots/fe1_residual_plots.png',
# width = 8, height = 3, units = 'in', res = 300) par(mfrow
# = c(1, 3))

# Residuals vs Fitted Values plot for fe_spi_di
plot(fe_spi_di, which = 1, main = "FE Residuals vs Fitted Model",
    caption = "fe_spi_di [Stage 1]", pch = 1, cex = 0.35, col = "#595959")
abline(h = 0, col = "black", lty = 2, lwd = 1)
fitted_vals_fe <- fitted(fe_spi_di)
lines(lowess(fitted_vals_fe, resid(fe_spi_di)), col = "red",
    lwd = 1.5)
```

FE Residuals vs Fitted Model

fe_spi_di [Stage 1]

Fitted values

lm(spi_comp ~ di_score + log_gdppc + factor(year) + factor(country_code))
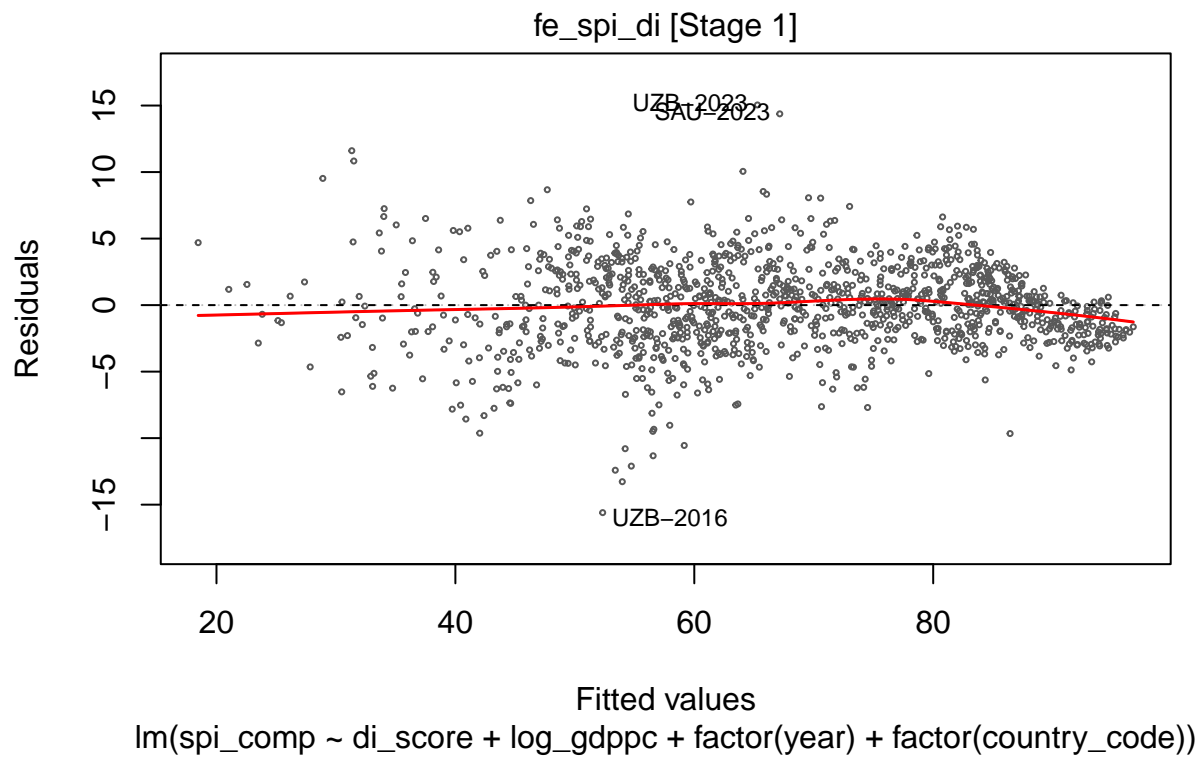
```r
# Residuals vs di_score plot
plot(model_data_fe1$di_score, resid(fe_spi_di), xlab = "di_score",
    ylab = "", yaxt = "n", main = "FE Residuals vs di_score",
    pch = 1, cex = 0.35, col = "darkred")
abline(h = 0, col = "black", lty = 2, lwd = 1)
lines(lowess(model_data_fe1$di_score, resid(fe_spi_di)), col = "red",
    lwd = 1.5)
```

FE Residuals vs di_score

di_score
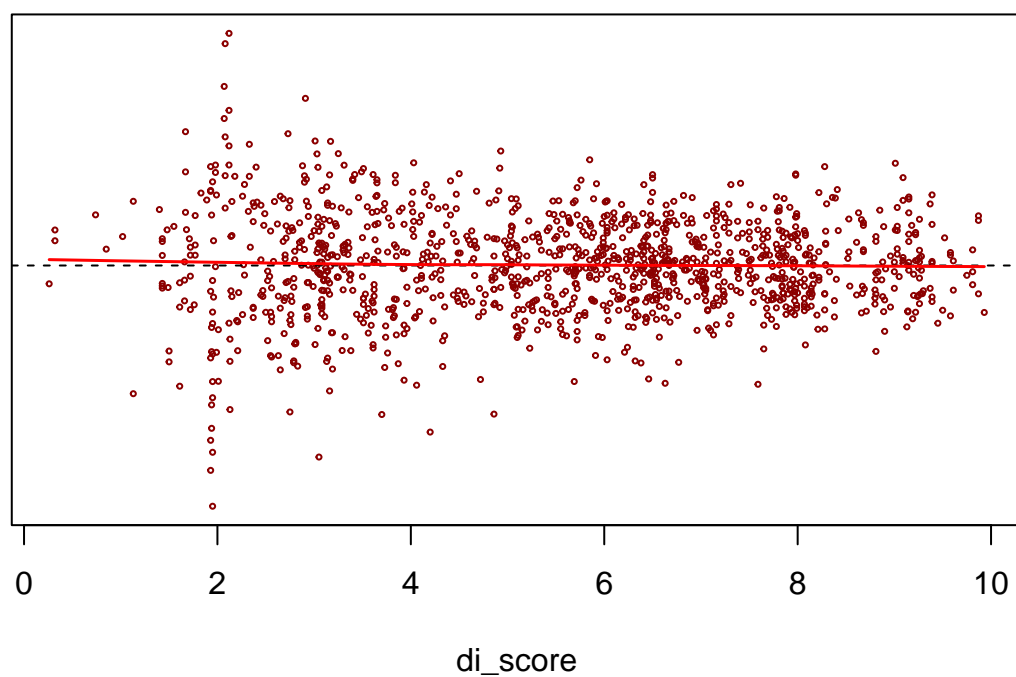
```r
# Residuals vs log_gdppc plot
plot(model_data_fe1$log_gdppc, resid(fe_spi_di), xlab = "log_gdppc",
    ylab = "", yaxt = "n", main = "FE Residuals vs log_gdppc",
    pch = 1, cex = 0.35, col = "darkgreen")
abline(h = 0, col = "black", lty = 2, lwd = 1)
lines(lowess(model_data_fe1$log_gdppc, resid(fe_spi_di)), col = "red",
    lwd = 1.5)
```

FE Residuals vs log_gdppc

log_gdppc
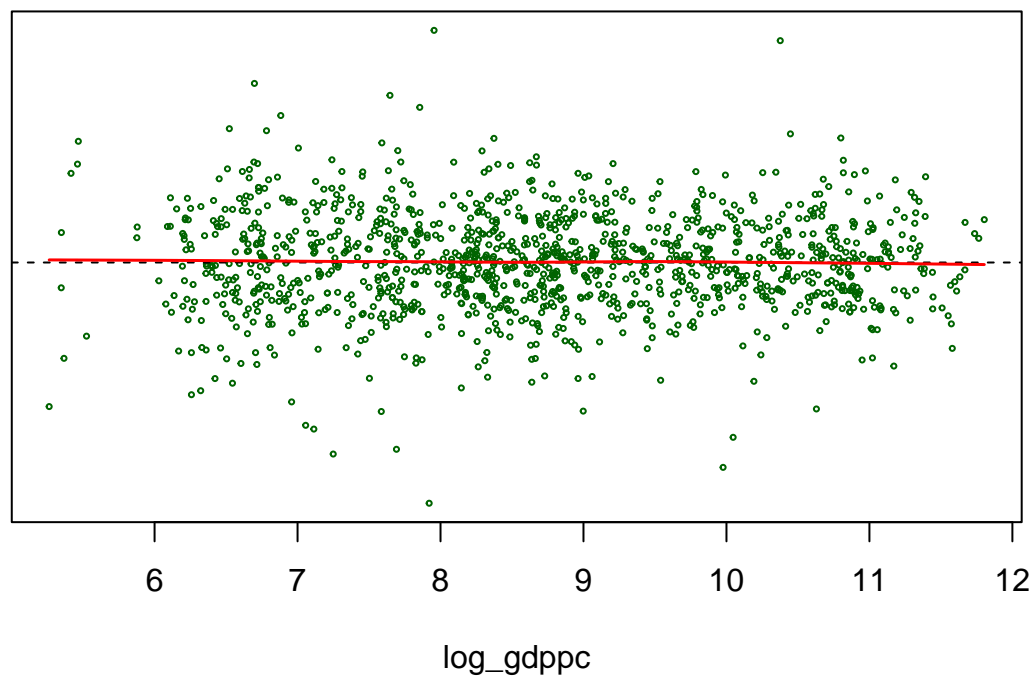
```r
dev.off() # to save
```

```
## quartz_off_screen
##                 3
```

```r
par(mfrow = c(1, 1))
```