Component 2: FE, FD, Granger Causality

Sevastian Sanchez

2025-05-05

#TEMPORAL MEDIATION ANALYSIS: First Difference, Fixed Effects, Lags, Checks

##setup: packages and data

TEST X: The effect of democracy score on SDG

```r
# Total Effect: Check if regime type directly affects SDG
# scores (without SPI) + Controls & RobustSEs
total_sdg_di_ols <- lm(sdg_overall ~ di_score + log_gdppc + population +
    year_fct, data = fd_fe_data)
summary(total_sdg_di_ols)
```

```
##
## Call:
## lm(formula = sdg_overall ~ di_score + log_gdppc + population +
##     year_fct, data = fd_fe_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.9534  -4.0034  -0.0776   3.9733  15.6468
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.747e+02  1.757e+02  -2.703  0.00699 **
## di_score     1.250e+00  1.060e-01  11.797  < 2e-16 ***
## log_gdppc    4.395e+00  1.603e-01  27.425  < 2e-16 ***
## population   2.521e-10  1.085e-09   0.232  0.81639
## year_fct     2.460e-01  8.704e-02   2.826  0.00479 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.757 on 1099 degrees of freedom
##   (240 observations deleted due to missingness)
## Multiple R-squared:  0.6884, Adjusted R-squared:  0.6872
## F-statistic: 606.9 on 4 and 1099 DF,  p-value: < 2.2e-16
```

```r
coeftest(total_sdg_di_ols, vcov = vcovHC(total_sdg_di_ols, type = "HC1"))
```

```
##
## t test of coefficients:
##
##                Estimate  Std. Error t value  Pr(>|t|)
## (Intercept) -4.7474e+02  1.7727e+02 -2.6780  0.007517 **
```

```
## di_score     1.2501e+00  1.3589e-01  9.1995 < 2.2e-16 ***
## log_gdppc    4.3952e+00  2.0052e-01 21.9194 < 2.2e-16 ***
## population   2.5207e-10  7.0258e-10  0.3588  0.719835
## year_fct     2.4600e-01  8.7843e-02  2.8005  0.005192 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# plot relationship
plot_sdg_di <- ggplot(fd_fe_data, aes(x = di_score, y = sdg_overall)) +
    geom_point(color = "#EE6A50", size = 1, alpha = 0.65) + geom_smooth(method = "lm",
    se = TRUE, color = "darkred", size = 1) + labs(title = "Effect of Democracy Score on SDG Status",
    x = "Democracy Score", y = "SDG Progress") + theme_minimal()
```

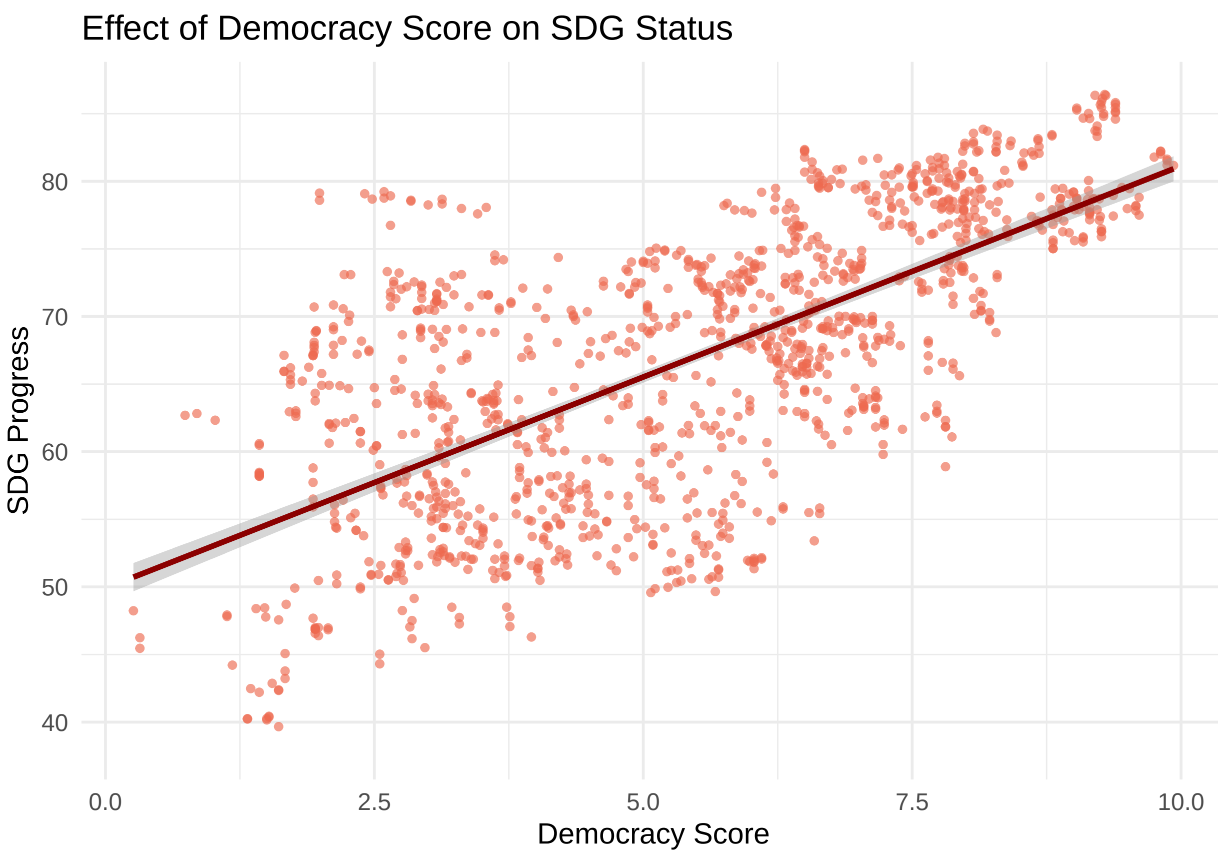
```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```r
plot_sdg_di
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 64 rows containing non-finite outside the scale range
## (`stat_smooth()`).
```

```
## Warning: Removed 64 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



TEST 1: OLS Mediation analysis [REDO RESULTS FOR ROBUST SEs]

To test if SPI mediates the relationship between regime type and SDG outcomes:
Democratic backsliding → reduces SPI → slows SDG progress

H0: SPI DOES NOT mediate (indirectly effect) the relationship between regime type and SDG status H1: SPI mediates (indirectly effects) the relationship between regime type and SDG status

- ACME (Average Causal Mediation Effect): SPI's indirect effect.
- ADE (Average Direct Effect): Regime type's direct effect, excluding SPI

```
# seed to reproduce estimates
set.seed(125)

# OLS Mediator model: Check if regime type affects SPI
med_spi_ols <- lm(spi_comp ~ di_score + log_gdppc + population +
    year_fct, data = fd_fe_data)
summary(med_spi_ols)
```

```
##
## Call:
## lm(formula = spi_comp ~ di_score + log_gdppc + population + year_fct,
##     data = fd_fe_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -33.666  -5.881   0.677   6.472  28.079
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.246e+03  3.236e+02 -13.121   <2e-16 ***
## di_score     3.689e+00  1.954e-01  18.880   <2e-16 ***
## log_gdppc    3.379e+00  2.941e-01  11.488   <2e-16 ***
## population   4.830e-09  1.980e-09   2.439   0.0149 *
## year_fct     2.111e+00  1.603e-01  13.168   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.49 on 1078 degrees of freedom
##   (261 observations deleted due to missingness)
## Multiple R-squared:  0.5906, Adjusted R-squared:  0.5891
## F-statistic: 388.9 on 4 and 1078 DF,  p-value: < 2.2e-16
```

```
coeftest(med_spi_ols, vcov = vcovHC(med_spi_ols, type = "HC1"))  #Robust SE
```

```
##
## t test of coefficients:
##
##                Estimate  Std. Error  t value  Pr(>|t|)
## (Intercept) -4.2460e+03  3.2755e+02 -12.9629 < 2.2e-16 ***
## di_score     3.6893e+00  2.2815e-01  16.1710 < 2.2e-16 ***
## log_gdppc    3.3787e+00  2.9147e-01  11.5922 < 2.2e-16 ***
## population   4.8298e-09  1.3743e-09   3.5144  0.000459 ***
## year_fct     2.1113e+00  1.6235e-01  13.0048 < 2.2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# OLS Outcome model: Check if SPI affects SDG scores while
# controlling for regime type
output_sdg_ols <- lm(sdg_overall ~ spi_comp + di_score + log_gdppc +
    population + year_fct, data = fd_fe_data)
summary(output_sdg_ols)
```

```
##
## Call:
## lm(formula = sdg_overall ~ spi_comp + di_score + log_gdppc +
##     population + year_fct, data = fd_fe_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.2515  -3.3258   0.0273   3.2133  13.3604
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.024e+02  1.597e+02   4.398 1.20e-05 ***
## spi_comp     2.864e-01  1.396e-02  20.522  < 2e-16 ***
## di_score     2.172e-01  1.033e-01   2.103   0.0357 *
## log_gdppc    3.300e+00  1.428e-01  23.112  < 2e-16 ***
## population  -1.215e-09  9.099e-10  -1.336   0.1820
## year_fct    -3.389e-01  7.916e-02  -4.281 2.02e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.808 on 1077 degrees of freedom
##   (261 observations deleted due to missingness)
## Multiple R-squared:  0.7736, Adjusted R-squared:  0.7725
## F-statistic: 735.8 on 5 and 1077 DF,  p-value: < 2.2e-16
```

```r
coeftest(output_sdg_ols, vcov = vcovHC(output_sdg_ols, type = "HC1"))  #Robust SE
```

```
##
## t test of coefficients:
##
##                Estimate  Std. Error t value  Pr(>|t|)
## (Intercept)  7.0238e+02  1.5950e+02  4.4036 1.171e-05 ***
## spi_comp     2.8641e-01  1.5627e-02 18.3280 < 2.2e-16 ***
## di_score     2.1725e-01  1.1547e-01  1.8814   0.06019 .
## log_gdppc    3.3001e+00  1.8662e-01 17.6836 < 2.2e-16 ***
## population  -1.2152e-09  8.4979e-10 -1.4300   0.15302
## year_fct    -3.3890e-01  7.9041e-02 -4.2877 1.967e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# OLS Mediation test: Quantify how much of regime type's
# effect on SDGs operates through SPI
med_model <- mediate(med_spi_ols, output_sdg_ols, treat = "di_score",
    mediator = "spi_comp", sims = 1000, robustSE = TRUE)
summary(med_model)
```

```
##
## Causal Mediation Analysis
```

```
##
## Quasi-Bayesian Confidence Intervals
##
##                  Estimate 95% CI Lower 95% CI Upper p-value
## ACME               1.0599       0.9015         1.24  <2e-16 ***
## ADE                0.2205      -0.0109         0.45   0.064 .
## Total Effect       1.2804       1.0118         1.55  <2e-16 ***
## Prop. Mediated     0.8255       0.6997         1.01  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Sample Size Used: 1083
##
##
## Simulations: 1000
```

```
# Sensitivity Analysis: Tests how robust mediation findings
# are to potential unmeasured confounding variables
# sens_med <- medsens(med_model, rho.by = 0.1)
# summary(sens_med) plot(sens_med)
```

[UPDATE RESULTS - OMITTED GINI VAR] ACEM: SPI's indirect effect = 0.62884 units ($p < 0.001$) ADE: Regime type's direct effect, excluding SPI = 0.00565 ($p = 0.008$) ??? units Total Effect: = 0.63449 units ($p = 0.008$) Proportion Mediated: = 0.98081 or 98.08% ?? of total units

`Proportion Mediated` = the percent difference between the total effect (SDG ~ DI) and indirect ACEM effect (+ SPI)

Interpretation: A 1-unit DI increase boosts SDG scores by 0.63449 total units, with 0.62884 units (98.08% of units) transmitted through SPI. The remaining 0.00565 units represents democracy score's effect (e.g., governance reforms unrelated to statistics) on SDG progress.

Because the indirect effect of SPI on sdg_overall (ACME) is highly significant ($p < 0.001$), there is evidence to suggest that SPI mediates the regime-SDG relationship, based on the model.

Because the ADE (the direct effect between di_score on sdg_overall) is not significant, SPI DOES ENTIRELY explain the connection between regime type and sdg status, based on the model.

Panel Data Analysis

First Difference (FD) removes all time-invariant characteristics of each unit (like geography, culture, or baseline wealth). However, variables that do change over time, such as GDP per capita and country population, should be controlled.

Fixed Effects (FE) [description…]

TEST X, First Difference: SPI ~ DI & SDG ~ SPI

```
# panel dataframe
panel_data <- pdata.frame(fd_fe_data, index = c("country_code",
    "year")  # Replace 'year' with your time variable
)
```

```
## Warning in pdata.frame(fd_fe_data, index = c("country_code", "year")): at least one NA in at least o
##  to find out which, use, e.g., table(index(your_pdataframe), useNA = "ifany")
```

```r
###### SPI ~ DI ######
fd_spi_di <- plm(spi_comp ~ di_score + log_gdppc + population,
    data = panel_data, model = "fd"  #FD
)
# summary(fd_spi_di) Country-level custard Robust SEs
coeftest(fd_spi_di, vcov = vcovHC(fd_spi_di, cluster = "group",
    type = "HC1"))
```

```
##
## t test of coefficients:
##
##                Estimate  Std. Error t value  Pr(>|t|)
## (Intercept)  1.9040e+00  1.0803e-01 17.6245 < 2.2e-16 ***
## di_score    -9.9274e-01  3.7520e-01 -2.6459 0.0082864 **
## log_gdppc    3.1951e+00  9.4261e-01  3.3896 0.0007295 ***
## population  -1.1730e-07  3.3133e-08 -3.5404 0.0004196 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
###### SDG ~ SPI ######
fd_sdg_spi <- plm(sdg_overall ~ spi_comp + di_score + log_gdppc +
    population, data = panel_data, model = "fd"  #FD
)
# summary(fd_sdg_spi) Country-level custard Robust SEs
coeftest(fd_sdg_spi, vcov = vcovHC(fd_sdg_spi, cluster = "group",
    type = "HC1"))
```

```
##
## t test of coefficients:
##
##               Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 2.9396e-01 1.7497e-02 16.8002 < 2.2e-16 ***
## spi_comp    1.5814e-02 5.3511e-03  2.9552 0.0032038 **
## di_score    7.1690e-02 8.0801e-02  0.8872 0.3751810
## log_gdppc   2.3069e-02 1.8497e-01  0.1247 0.9007737
## population  2.5005e-08 6.5546e-09  3.8149 0.0001453 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results:

FD Models: SPI ~ DI: -0.9927 (p = 0.0083) **SDG ~ SPI: 0.0158 (p = 0.0032)**

TEST X, Fixed Effects Models: SPI ~ DI & SDG ~ SPI

```r
###### SPI ~ DI ######
fe_spi_di <- plm(spi_comp ~ di_score + log_gdppc + population +
    factor(year), data = panel_data, model = "within"  #FE
)

summary(fe_spi_di)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = spi_comp ~ di_score + log_gdppc + population +
```

```
##     factor(year), data = panel_data, model = "within")
##
## Unbalanced Panel: n = 156, T = 1-7, N = 1083
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -13.519427  -1.995078   0.019582   2.039475  12.247203
##
## Coefficients:
##                     Estimate  Std. Error t-value  Pr(>|t|)
## di_score         -2.3840e-01  3.6906e-01 -0.6460   0.51846
## log_gdppc         7.0828e-01  1.0536e+00  0.6723   0.50159
## population       -7.1978e-08  3.5720e-08 -2.0151   0.04419 *
## factor(year)2017  2.5684e+00  3.9460e-01  6.5089 1.244e-10 ***
## factor(year)2018  5.0431e+00  4.0955e-01 12.3139 < 2.2e-16 ***
## factor(year)2019  5.3559e+00  4.1166e-01 13.0105 < 2.2e-16 ***
## factor(year)2020  7.7324e+00  4.0068e-01 19.2983 < 2.2e-16 ***
## factor(year)2021  1.2490e+01  4.4430e-01 28.1129 < 2.2e-16 ***
## factor(year)2022  1.1993e+01  4.8042e-01 24.9647 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    30321
## Residual Sum of Squares: 10577
## R-Squared:      0.65115
## Adj. R-Squared: 0.58883
## F-statistic: 190.392 on 9 and 918 DF, p-value: < 2.22e-16
```

```r
coeftest(fe_spi_di, vcov = vcovHC(fe_spi_di, cluster = "group",
    type = "HC1"))  # Robust SEs
```

```
##
## t test of coefficients:
##
##                     Estimate  Std. Error t value Pr(>|t|)
## di_score         -2.3840e-01  5.7563e-01 -0.4142   0.6789
## log_gdppc         7.0828e-01  2.0063e+00  0.3530   0.7242
## population       -7.1978e-08  5.1197e-08 -1.4059   0.1601
## factor(year)2017  2.5684e+00  2.7661e-01  9.2854   <2e-16 ***
## factor(year)2018  5.0431e+00  4.7914e-01 10.5254   <2e-16 ***
## factor(year)2019  5.3559e+00  5.3354e-01 10.0385   <2e-16 ***
## factor(year)2020  7.7324e+00  5.5680e-01 13.8873   <2e-16 ***
## factor(year)2021  1.2490e+01  7.7008e-01 16.2197   <2e-16 ***
## factor(year)2022  1.1993e+01  8.6247e-01 13.9059   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
##### SDG ~ SPI #####
fe_sdg_spi <- plm(sdg_overall ~ spi_comp + di_score + log_gdppc +
    population + factor(year), data = panel_data, model = "within"   #FE
)

summary(fe_sdg_spi)
```

```
## Oneway (individual) effect Within Model
```

```
##
## Call:
## plm(formula = sdg_overall ~ spi_comp + di_score + log_gdppc +
##     population + factor(year), data = panel_data, model = "within")
##
## Unbalanced Panel: n = 156, T = 1-7, N = 1083
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -2.4556849 -0.2859821  0.0019399  0.2979312  2.6452072
##
## Coefficients:
##                    Estimate Std. Error t-value  Pr(>|t|)
## spi_comp         3.6294e-02 5.7386e-03  6.3245 3.966e-10 ***
## di_score         1.4105e-01 6.4184e-02  2.1975   0.02823 *
## log_gdppc        8.5293e-01 1.8323e-01  4.6549 3.720e-06 ***
## population       2.6318e-08 6.2244e-09  4.2282 2.592e-05 ***
## factor(year)2017 3.1120e-01 7.0175e-02  4.4346 1.034e-05 ***
## factor(year)2018 5.1169e-01 7.6865e-02  6.6570 4.807e-11 ***
## factor(year)2019 8.9149e-01 7.7897e-02 11.4445 < 2.2e-16 ***
## factor(year)2020 1.2561e+00 8.2598e-02 15.2078 < 2.2e-16 ***
## factor(year)2021 1.2029e+00 1.0538e-01 11.4143 < 2.2e-16 ***
## factor(year)2022 1.3474e+00 1.0823e-01 12.4490 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    864.8
## Residual Sum of Squares: 319.42
## R-Squared:      0.63065
## Adj. R-Squared: 0.56419
## F-statistic: 156.573 on 10 and 917 DF, p-value: < 2.22e-16
```

```r
coeftest(fe_sdg_spi, vcov = vcovHC(fe_sdg_spi, cluster = "group",
    type = "HC1"))  # Robust SEs
```

```
##
## t test of coefficients:
##
##                    Estimate Std. Error t value  Pr(>|t|)
## spi_comp         3.6294e-02 1.3288e-02  2.7313 0.0064296 **
## di_score         1.4105e-01 1.0676e-01  1.3212 0.1867630
## log_gdppc        8.5293e-01 2.7287e-01  3.1258 0.0018292 **
## population       2.6318e-08 7.4239e-09  3.5450 0.0004124 ***
## factor(year)2017 3.1120e-01 4.9978e-02  6.2267 7.242e-10 ***
## factor(year)2018 5.1169e-01 8.3517e-02  6.1268 1.330e-09 ***
## factor(year)2019 8.9149e-01 8.8693e-02 10.0514 < 2.2e-16 ***
## factor(year)2020 1.2561e+00 9.6461e-02 13.0221 < 2.2e-16 ***
## factor(year)2021 1.2029e+00 1.4450e-01  8.3245 3.056e-16 ***
## factor(year)2022 1.3474e+00 1.4657e-01  9.1929 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results:

SPI ~ DI: -0.4976 (p = 0.3886) [not significant] year_fct: 2.0536 (p < 2e-16***) SDG ~ SPI: 0.0324 (p = 0.1247) [not significant] year_fct: 2.5292e-01 (p < 2.2e-16***)***

Neither SPI~DI or SDG~SPI models are statistically significant. This contrasts from previous first difference estimates that found both SPI~DI and SDG~SPI models statistically significant ($p < 0.05$). Interestingly, the factored-year variable (year_fct) in both FE models are extremely statistically significant with increasing coefficient estimates over the years. This suggests underlying trends that statistical capacity is, on average, increasing based on the model.

TEST X: Granger Causality: DI & SPI

Essentially tests whether a predictor variable in a prior year has an affect on a present day dependent variable. In this case, my study utilizes a granger test to both understand directionality of effects, assessing potential reverse causality of x and y variables, and determine the need to incorporate lagged predictors (DI and SPI variables).

```r
# SPI ~ DI [SIGNIFICANT]
grangertest(spi_comp ~ di_score, order = 1, data = panel_data)  # Most optimal
```

```
## Granger causality test
##
## Model 1: spi_comp ~ Lags(spi_comp, 1:1) + Lags(di_score, 1:1)
## Model 2: spi_comp ~ Lags(spi_comp, 1:1)
##   Res.Df Df     F    Pr(>F)
## 1   1243
## 2   1244 -1 16.97 4.048e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
grangertest(spi_comp ~ di_score, order = 2, data = panel_data)
```

```
## Granger causality test
##
## Model 1: spi_comp ~ Lags(spi_comp, 1:2) + Lags(di_score, 1:2)
## Model 2: spi_comp ~ Lags(spi_comp, 1:2)
##   Res.Df Df      F  Pr(>F)
## 1   1240
## 2   1242 -2 8.8021 0.00016 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# DI ~ SPI [NOT]
grangertest(di_score ~ spi_comp, order = 1, data = panel_data)
```

```
## Granger causality test
##
## Model 1: di_score ~ Lags(di_score, 1:1) + Lags(spi_comp, 1:1)
## Model 2: di_score ~ Lags(di_score, 1:1)
##   Res.Df Df      F Pr(>F)
## 1   1243
## 2   1244 -1 0.4086 0.5228
```

```r
grangertest(di_score ~ spi_comp, order = 2, data = panel_data)
```

```
## Granger causality test
##
## Model 1: di_score ~ Lags(di_score, 1:2) + Lags(spi_comp, 1:2)
## Model 2: di_score ~ Lags(di_score, 1:2)
##   Res.Df Df      F Pr(>F)
## 1   1240
```

```
## 2   1242 -2 0.2382 0.7881
```

spi_comp ~ di_score[order=1]: F = 16.97, p = 4.048e-05 (***) - As speculated/hypothesized, democracy scores (di_score) DO help predict current SPI scores. - Although we recieve statistically significant results upon increasing the lag order, the F statistic drops sharply from 16.97 to 8.8021 and then continues falling gradually, as p-value increases ever so slightly (although still extremely below alpha 0.05). Accordingly, 1st orderd lagged di_score predictors are optimal and sufficient for the following Fixed Effects models.

di_score ~ spi_comp[order=1]: F = 0.4086, p = 0.5228 - As expected, past SPI scores do NOT help predict current democracy scores (di_score) - even incorporating more lags, the relationships remain not statistically significant

Bottom Line Changes in regime characteristics precede changes in statistical capacity, NOT vice versa

TEST X: Granger Causality: SPI & SDG

```r
# SDG ~ SPI [SIGNIFICANT?]
grangertest(sdg_overall ~ spi_comp, order = 1, data = panel_data)
```

```
## Granger causality test
##
## Model 1: sdg_overall ~ Lags(sdg_overall, 1:1) + Lags(spi_comp, 1:1)
## Model 2: sdg_overall ~ Lags(sdg_overall, 1:1)
##   Res.Df Df      F Pr(>F)
## 1   1296
## 2   1297 -1 2.5447 0.1109
```

```r
grangertest(sdg_overall ~ spi_comp, order = 2, data = panel_data)
```

```
## Granger causality test
##
## Model 1: sdg_overall ~ Lags(sdg_overall, 1:2) + Lags(spi_comp, 1:2)
## Model 2: sdg_overall ~ Lags(sdg_overall, 1:2)
##   Res.Df Df      F Pr(>F)
## 1   1293
## 2   1295 -2 1.8186 0.1627
```

```r
# SPI ~ SDG [NOT?]
grangertest(spi_comp ~ sdg_overall, order = 1, data = panel_data)  # most optimal
```

```
## Granger causality test
##
## Model 1: spi_comp ~ Lags(spi_comp, 1:1) + Lags(sdg_overall, 1:1)
## Model 2: spi_comp ~ Lags(spi_comp, 1:1)
##   Res.Df Df      F  Pr(>F)
## 1   1296
## 2   1297 -1 8.7453 0.00316 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
grangertest(spi_comp ~ sdg_overall, order = 2, data = panel_data)
```

```
## Granger causality test
##
## Model 1: spi_comp ~ Lags(spi_comp, 1:2) + Lags(sdg_overall, 1:2)
## Model 2: spi_comp ~ Lags(spi_comp, 1:2)
##   Res.Df Df      F Pr(>F)
```

```
## 1   1293
## 2   1295 -2 4.5824 0.0104 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results:

SDG~SPI [order = 1]: F = 2.5447, p = 0.1109

SPI~SDG [order = 1]: F = 8.7453, p = 0.00316 (**)

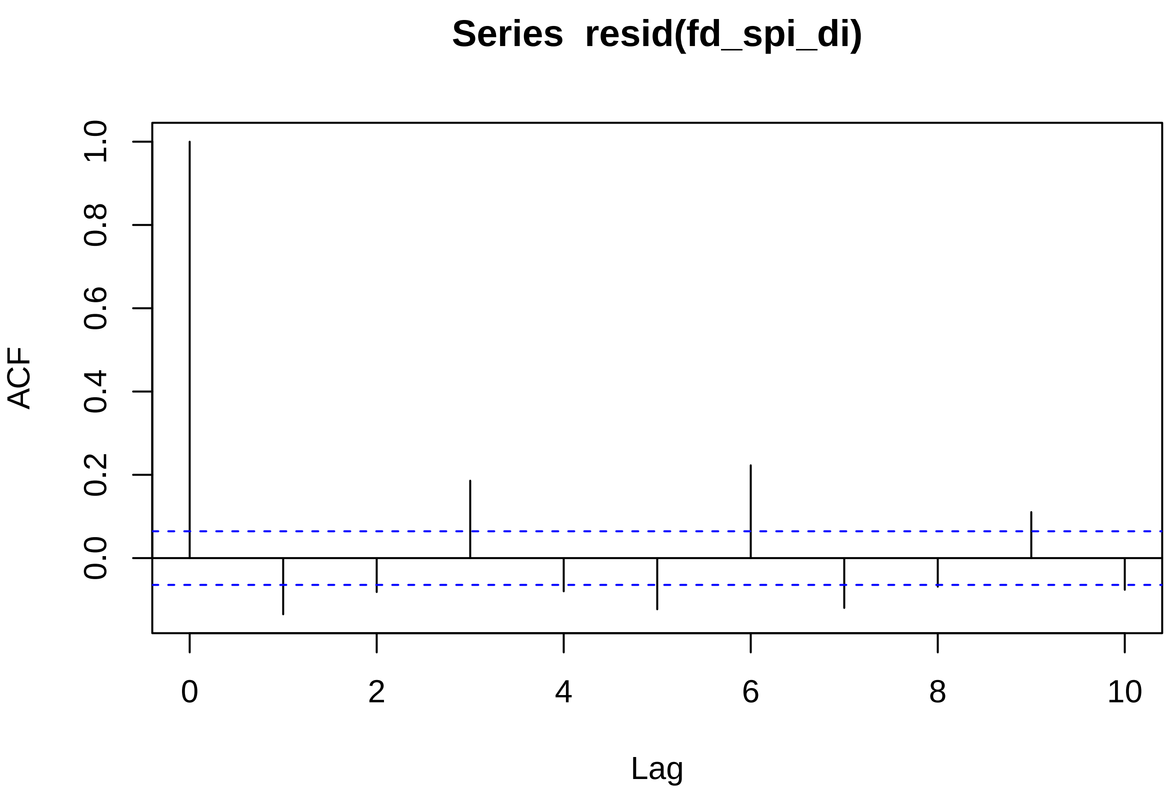
Unexpectedly, the results of the granger causality tests showed that it is SDG status that preceeds changes in statistical capacity, similar to DI score. It is possible that certain SDG goals/indicators could predict statistical capacity, such as SDG 17 (strong institutions) and/or SDG 4 (Quality Education) that could facilitate the growth of national statistical capacity. Deeper analysis of separate SDG score (disaggregated by individual score 1-17) impacts on predictor variables (in a similar methodological scheme) would provide interesting insights into this 'chicken or the egg' problem on a deeper level. This study reconsiders strategy due to suprising, but valid, evidence of reverse causality.

Assessing Autocorrelation

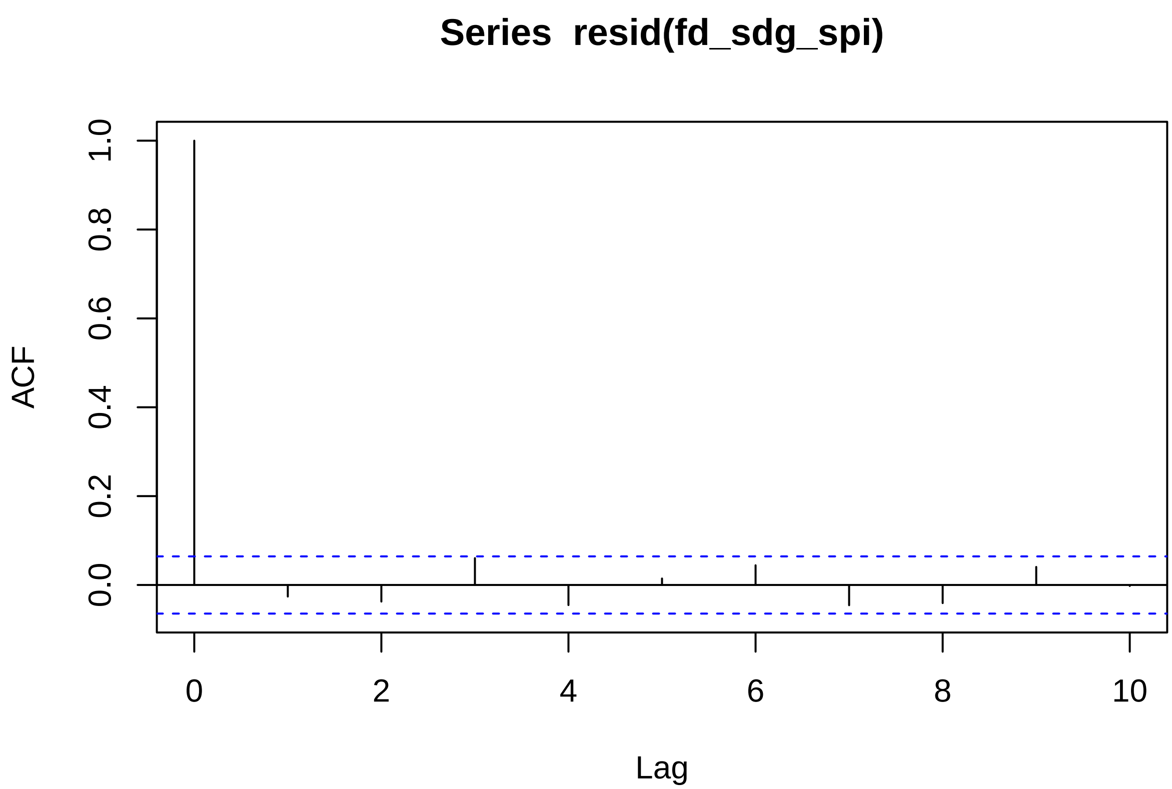
ACF checks for independence of errors in regression analysis. This assesses whether x variables are correlated across time to determine correlation at different lags (autocorrelation).

ACF Function

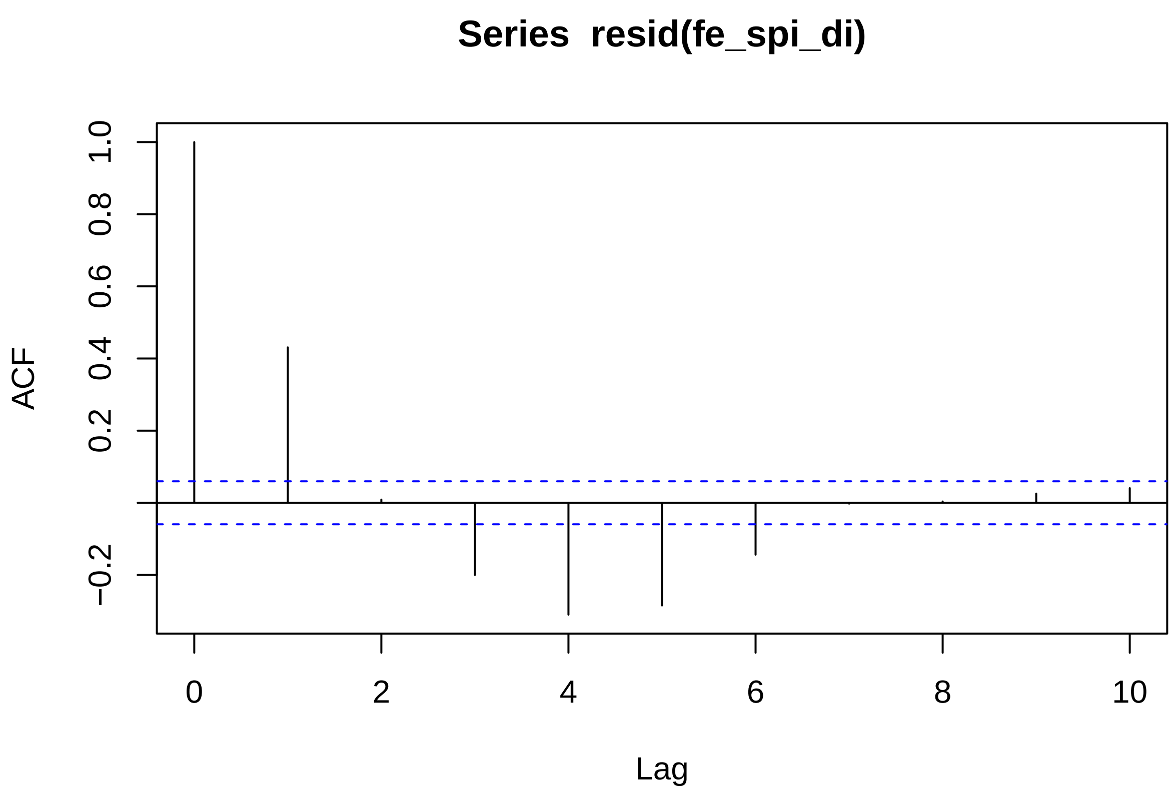
```r
acf(resid(fd_spi_di), lag.max = 10)
```



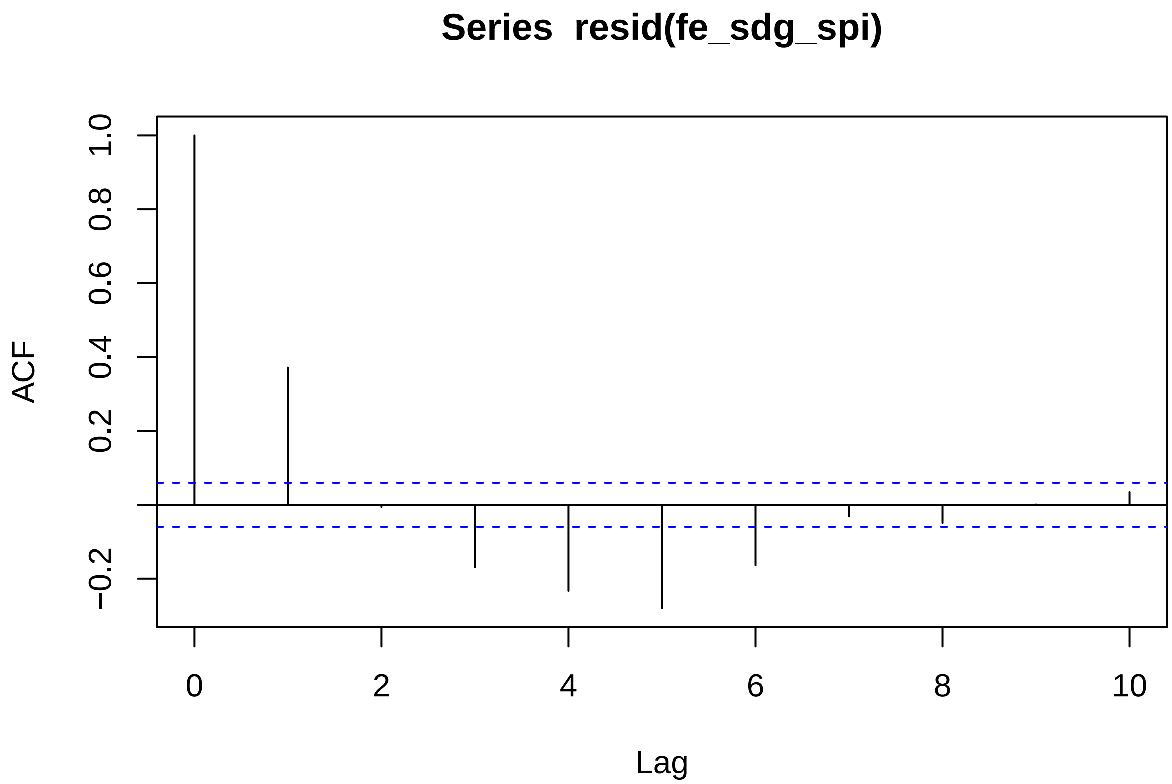

```r
acf(resid(fd_sdg_spi), lag.max = 10)
```




```r
acf(resid(fe_spi_di), lag.max = 10)
```




```r
acf(resid(fe_sdg_spi), lag.max = 10)
```



Temporal Dependent Lags

Considering results from the granger test, there is evidence that democracy score changes preceed changes in SPI (first). This is in line with studies that sustain it can take some time for implementation of statistical capacity structures to produce results.

Similarly, I suspect that regime transition has a delayed effect on statistical capacity, although I suspect it varies in time (do we notice a drop in SPI as a precondition or a symptom of autocratization)?

Significant Autocorrelation: create lagged variables and first differences [MAY NOT NEED]

TEST X, First Difference + LAGS: SPI ~ DI & SDG ~ SPI [R STUDIO CRASHES HERE-BEYOND THIS POINT]

TEST X, Fixed Effects + LAGS: SPI ~ DI & SDG ~ SPI

ROBUSTNESS CHECKS

AIC/BIC

Breusch-Godfrey Autocorrelation Tests: AR(1) + AR(2)

```
# mediator models
pbgtest(fe_spi_di, order = 1)  # AR(1)
```

```
##
##  Breusch-Godfrey/Wooldridge test for serial correlation in panel models
##
## data:  spi_comp ~ di_score + log_gdppc + population + factor(year)
```

```
## chisq = 201.56, df = 1, p-value < 2.2e-16
## alternative hypothesis: serial correlation in idiosyncratic errors
```

```r
pbgtest(fe_spi_di, order = 2)  # AR(2)
```

```
##
##  Breusch-Godfrey/Wooldridge test for serial correlation in panel models
##
## data:  spi_comp ~ di_score + log_gdppc + population + factor(year)
## chisq = 243.01, df = 2, p-value < 2.2e-16
## alternative hypothesis: serial correlation in idiosyncratic errors
```

This analysis does not test for autocorrelation by using the Durban Watson approach because the approach is designed for single time series, not panel models. Instead, this study adopts the Breusch-Godfrey (BG) approach to detect *high order* autocorrelation in addition to other forms.

Mediation Effect: connecting Regime -> SDGs

Mediation analysis to formally test and quantify SPI's role in transmitting regime change effects to SDGs. This is how to quantify the connection between the Regime Change and the SDGs

Mediation Effect + Lags: connecting DI -> SDGs

FD Mediation (including lags) [MAY NOT NEED]

FE Mediation (including lags) [MAY NOT NEED]