

Wrangle Marius

Marius Cotorobai

```
data <- read.csv(file = 'marius.csv')
colnames(data) <- c("country_code",
                    "E", "GDP",
                    "A", "I", "S")
data$E <- as.numeric(data$E)
data$GDP <- as.numeric(data$GDP)
data$A <- as.numeric(data$A)
data$I <- as.numeric(data$I)
data$S <- as.numeric(data$S)

# Since you have all different countries, there is no point in replacing
# country codes with numbers from 1 to 121

write.csv(data, file = "cleaned_marius.csv")

# Please note that you don't need S because it will be captured as the base
# effect when both of your A and I will be 0. Please refer to the next code
# chunk in order to see a better way to run this.

cleaned_data <- read.csv(file = "cleaned_marius.csv")
naive_ols <- lm(E ~ GDP*A + GDP*I + GDP*S + I(GDP^2)*A + I(GDP^2)*I + I(GDP^2)*S,
               data = cleaned_data)
summary(naive_ols)

##
## Call:
## lm(formula = E ~ GDP * A + GDP * I + GDP * S + I(GDP^2) * A +
##     I(GDP^2) * I + I(GDP^2) * S, data = cleaned_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.33859 -0.08394 -0.02119  0.04756  0.60263
##
## Coefficients: (3 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.302e-01  6.362e-02   3.619 0.000446 ***
## GDP          5.633e-06  4.538e-06   1.241 0.217115
## A           -2.139e-01  7.894e-02  -2.709 0.007806 **
## I            3.162e-01  1.063e-01   2.976 0.003584 **
## S              NA           NA      NA      NA
## I(GDP^2)     -7.259e-11  6.660e-11  -1.090 0.278028
## GDP:A         6.360e-05  1.684e-05   3.776 0.000257 ***
## GDP:I        -9.133e-06  8.636e-06  -1.058 0.292522
## GDP:S              NA           NA      NA      NA
```

```
## A:I(GDP^2) -2.175e-09 7.300e-10 -2.980 0.003534 **
## I:I(GDP^2) 6.820e-11 1.113e-10 0.613 0.541380
## S:I(GDP^2) NA NA NA NA
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1692 on 112 degrees of freedom
## Multiple R-squared: 0.3855, Adjusted R-squared: 0.3416
## F-statistic: 8.784 on 8 and 112 DF, p-value: 2.905e-09
```

```
cleaned_data2 <- read.csv(file = "cleaned_marius.csv")
ols <- lm(E ~ GDP*A + GDP*I + I(GDP^2)*A + I(GDP^2)*I, data = cleaned_data2)
summary(ols)
```

```
##
## Call:
## lm(formula = E ~ GDP * A + GDP * I + I(GDP^2) * A + I(GDP^2) *
##     I, data = cleaned_data2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.33859 -0.08394 -0.02119  0.04756  0.60263
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.302e-01  6.362e-02   3.619 0.000446 ***
## GDP          5.633e-06  4.538e-06   1.241 0.217115
## A           -2.139e-01  7.894e-02  -2.709 0.007806 **
## I            3.162e-01  1.063e-01   2.976 0.003584 **
## I(GDP^2)     -7.259e-11  6.660e-11  -1.090 0.278028
## GDP:A         6.360e-05  1.684e-05   3.776 0.000257 ***
## GDP:I        -9.133e-06  8.636e-06  -1.058 0.292522
## A:I(GDP^2)   -2.175e-09  7.300e-10  -2.980 0.003534 **
## I:I(GDP^2)    6.820e-11  1.113e-10   0.613 0.541380
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1692 on 112 degrees of freedom
## Multiple R-squared: 0.3855, Adjusted R-squared: 0.3416
## F-statistic: 8.784 on 8 and 112 DF, p-value: 2.905e-09
```