Dhyey Mavani Code for ECON-361 Final Empirical Project

Dhyey Mavani

15 December 2022

```
data <- read.csv(file = 'imported_data.csv')
colnames(data) <- c("year", "year_code", "country_name", "country_code",
                    "pct_GDP_val_add_by_serv", "pct_urban_popl",
                    "pct_GDP_govt_exp_on_edu")
data <- subset(data, select = -c(year_code,country_name))
data$year <- as.numeric(data$year)
```

```
## Warning: NAs introduced by coercion
```

```
data$pct_GDP_govt_exp_on_edu <- as.numeric(data$pct_GDP_govt_exp_on_edu)
```

```
## Warning: NAs introduced by coercion
```

```
data$pct_GDP_val_add_by_serv <- as.numeric(data$pct_GDP_val_add_by_serv)
data$country_code[data$country_code == c("AUS","AUT","BEL","CAN","CHL","CZE",
                                         "DNK","EST","FIN","FRA","DEU","GRC",
                                         "HUN","ISL","IRL","ISR","ITA","JPN",
                                         "KOR","LVA","LTU","LUX","MEX","NLD",
                                         "NZL","NOR","POL","PRT","SVK","SVN",
                                         "ESP","SWE","CHE","TUR","GBR","USA")] <-
  c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,
     19,20,21,22,23,24,25,26,27,28,29,30,31,31,33,34,35,36)
```

```
## Warning in data$country_code == c("AUS", "AUT", "BEL", "CAN", "CHL", "CZE", :
## longer object length is not a multiple of shorter object length
```

```
data$country_code <- as.numeric(data$country_code)
```

```
## Warning: NAs introduced by coercion
```

```
data <- tidyr::drop_na(data)
write.csv(data, file = "dhyey_cleaned_data.csv")
```

```
cleaned_data <- read.csv(file = "dhyey_cleaned_data.csv")
# install.packages("ivreg")
library("ivreg")
iv1 <- ivreg(pct_urban_popl ~ pct_GDP_val_add_by_serv | pct_GDP_govt_exp_on_edu, data = cleaned_data)
summary(iv1)
```

```
##
## Call:
## ivreg(formula = pct_urban_popl ~ pct_GDP_val_add_by_serv | pct_GDP_govt_exp_on_edu,
##     data = cleaned_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -176.818  -37.138   -3.393   43.854  148.975
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -566.124    534.090  -1.060    0.290
## pct_GDP_val_add_by_serv   10.410      8.641   1.205    0.229
##
## Diagnostic tests:
##                  df1 df2 statistic  p-value
## Weak instruments   1 304     1.306    0.254
## Wu-Hausman         1 303    33.816 1.54e-08 ***
## Sargan             0  NA        NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 52.31 on 304 degrees of freedom
## Multiple R-Squared: -20.29,  Adjusted R-squared: -20.36
## Wald test: 1.451 on 1 and 304 DF,  p-value: 0.2292
```

```
naive_ols <- lm(pct_urban_popl ~ pct_GDP_val_add_by_serv, data = cleaned_data)
summary(naive_ols)
```

```
##
## Call:
## lm(formula = pct_urban_popl ~ pct_GDP_val_add_by_serv, data = cleaned_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.5568  -8.5008   0.2333   9.0351  19.6942
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               30.3972     7.1043   4.279 2.52e-05 ***
## pct_GDP_val_add_by_serv   0.7591     0.1145   6.628 1.55e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.6 on 304 degrees of freedom
## Multiple R-squared:  0.1263, Adjusted R-squared:  0.1234
## F-statistic: 43.94 on 1 and 304 DF,  p-value: 1.545e-10
```

```
stage1 <- lm(pct_GDP_val_add_by_serv ~ pct_GDP_govt_exp_on_edu, data = cleaned_data)
summary(stage1)
```

```
##
## Call:
```

```
## lm(formula = pct_GDP_val_add_by_serv ~ pct_GDP_govt_exp_on_edu,
##     data = cleaned_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.7917  -4.6825   0.0958   3.6448  18.7464
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               60.3572     1.3048  46.257   <2e-16 ***
## pct_GDP_govt_exp_on_edu   0.2804     0.2453   1.143    0.254
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.296 on 304 degrees of freedom
## Multiple R-squared:  0.004279,   Adjusted R-squared:  0.001004
## F-statistic: 1.306 on 1 and 304 DF,  p-value: 0.2539
```

```
cleaned_data[["pred_pct_GDP_val_add_by_serv"]] <- predict(stage1, cleaned_data)
stage2 <- lm(pct_urban_popl ~ pred_pct_GDP_val_add_by_serv, data = cleaned_data)
summary(stage2)
```

```
##
## Call:
## lm(formula = pct_urban_popl ~ pred_pct_GDP_val_add_by_serv, data = cleaned_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.4524  -8.6490   0.5875   9.4510  20.1842
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   -566.124    109.724  -5.160 4.48e-07 ***
## pred_pct_GDP_val_add_by_serv   10.410      1.775   5.864 1.18e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.75 on 304 degrees of freedom
## Multiple R-squared:  0.1016, Adjusted R-squared:  0.09867
## F-statistic: 34.39 on 1 and 304 DF,  p-value: 1.176e-08
```