Midterm

Jonathan Eng

April 8, 2020

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(stargazer)
```

```
##
## Please cite as:

##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.

##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
library(ggplot2)
```

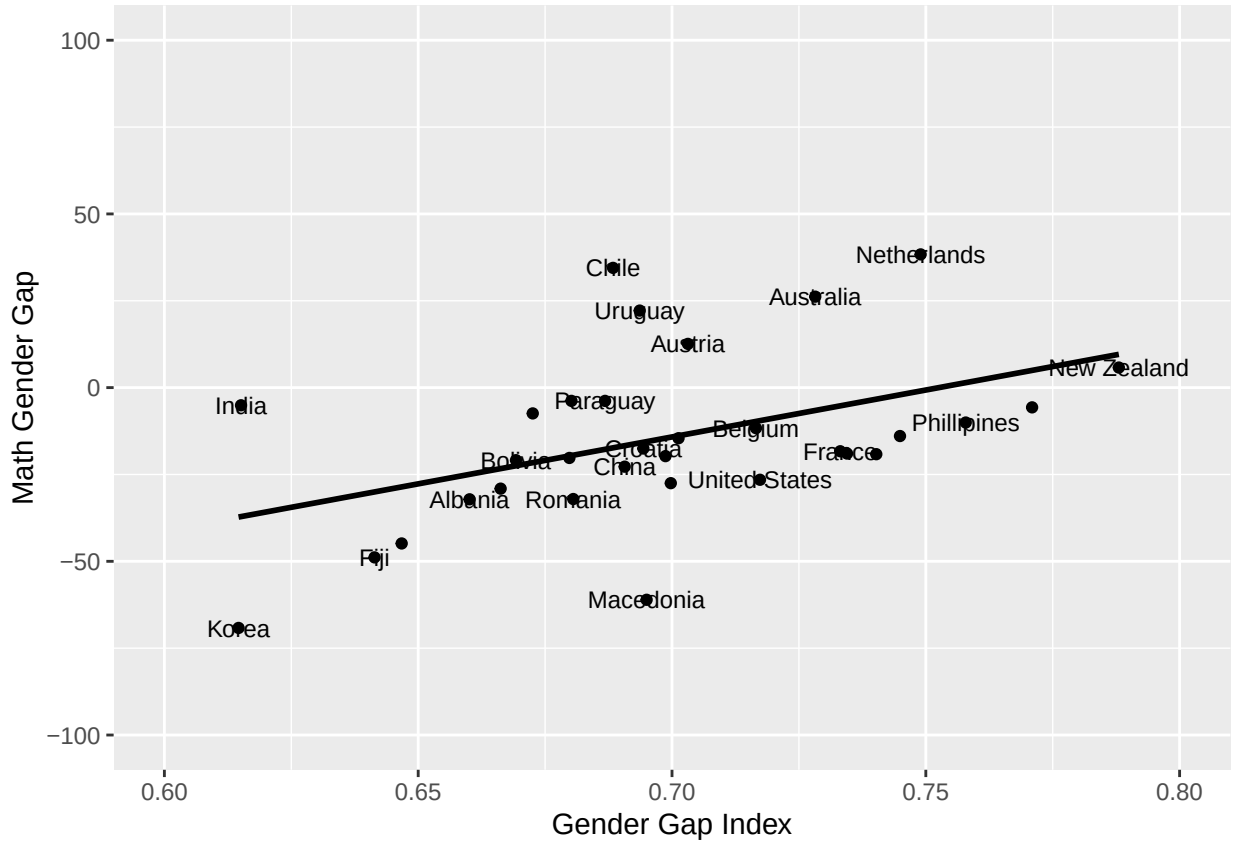
```
#Import MathData
mathData = read.csv("Final_sample.csv")
mathData = data.frame(mathData)
```

```
#Model Math Gender Gap
ols_model <- function(d){
  y <- d$pv1math
  x <- d$female
  tmp <- lm(y ~ x)
  return(tmp$coefficients[2])
}

mathGenderGap = mathData %>%
  group_by(background) %>%
```

```
  group_map(~ols_model(.x))

mathGenderGap <- unlist(mathGenderGap)
mathGenderGap <- unname(mathGenderGap)
```

```
#Plot Math Gender Gap
mathData %>%
  group_by(background) %>%
  summarise(genderGap = mean(ggi)) %>%
  ggplot(aes(x = genderGap, y = mathGenderGap)) +
    geom_point(size = 1.5, color = "black") +
    xlab("Gender Gap Index") +
    ylab("Math Gender Gap") +
    geom_text(size = 3, color = "black", check_overlap = TRUE, aes(label = background)) +
    ylim(-100, 100) +
    xlim(.6, .8) +
    geom_smooth(method = "lm", se = FALSE, color = "black")
```

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




```
# Note that in all specifications the dependent variables is pv1math.
# You need to include year fixed effects ( t),
# ancestry country fixed effects ( j),
# host country fixed effects ( k),
```

```
# and the interaction of female dummy with host country fixed effects (female_i × k)

#Model 1: Required variables above
model1 = lm(pv1math ~ female + I(ggi*female) + age + I(age*female) + diffgrade + I(diffgrade*female) + i
                      factor(background) + #( j)
                      factor(country) + (female*factor(country)), #( k)

                      data = mathData, weights = stweight)
model1$coefficients = model1$coefficients[1:6]
model1$coefficients
```

```
##     (Intercept)          female I(ggi * female)             age I(age * female)
##      370.278661     -164.648347      109.162276        4.230323        5.358375
##       diffgrade
##      -14.103388
```

```
#Model 2: Includes LogGDP [to model1]
model2 = lm(pv1math ~ female + I(ggi*female) + age + I(age*female) + diffgrade + I(diffgrade*female) + i
                      factor(background) + #( j)
                      factor(country) + (female*factor(country)) + #( k)
                      I(lgdppc*female), #(GDP * Female)

                      data = mathData, weights = stweight)
model2$coefficients = model2$coefficients[c(1:6,53)]
model2$coefficients
```

```
##        (Intercept)             female    I(ggi * female)                 age
##         365.991380        -154.427092         154.470319           4.374343
##    I(age * female)           diffgrade I(lgdppc * female)
##           5.187458         -14.255006          -4.575090
```

```
#Model 3: Removed ancestry country fixed effects ( j) [from model 2], includes gdp andggi and gdp
model3 = lm(pv1math ~ female + I(ggi*female) + age + I(age*female) + diffgrade + I(diffgrade*female) + i
                      factor(country) + female*factor(country) + #( k)
                      I(lgdppc*female) + #(GDP * Female)
                      ggi + lgdppc, #Gender Gap Index & Log per capital GDP

                      data = mathData, weights = stweight)
model3$coefficients = model3$coefficients[c(1:6,19:21)]
model3$coefficients
```

```
##        (Intercept)             female    I(ggi * female)                 age
##         175.862077         -86.357884         147.085635           5.934222
##    I(age * female)           diffgrade I(lgdppc * female)                ggi
##           1.472159         -17.148056          -5.203002         103.662942
##             lgdppc
##           3.669656
```

```
#Model 4: Included Parental Influence (Parental Education & Work) [to model 2]
model4 = lm(pv1math ~ female + I(ggi*female) + age + I(age*female) + diffgrade + I(diffgrade*female) + i
                      factor(background) + #( j)
```

```
                    factor(country) + (female*factor(country)) + #( k)
                    I(lgdppc*female) + #(GDP * Female)
                    fisced + I(fisced*female) + misced + I(misced*female) + dadwork + I(dadwork*female

                    data = mathData, weights = stweight)
model4$coefficients = model4$coefficients[c(1:6,53:61)]
model4$coefficients
```

```
##         (Intercept)              female     I(ggi * female)                 age
##         280.8046280         -99.8756533         179.8031588           6.1418717
##     I(age * female)            diffgrade  I(lgdppc * female)              fisced
##           0.2340563         -10.7840429          -2.9305991           6.8651007
##  I(fisced * female)              misced  I(misced * female)             dadwork
##          -2.1654029           2.9003971           0.3069656          20.7020324
## I(dadwork * female)             momwork I(momwork * female)
##          -3.7253272          15.0424259         -12.1484799
```

```
#Model 5: Included Home Possessions [to model 4]
model5 = lm(pv1math ~ female + I(ggi*female) + age + I(age*female) + diffgrade + I(diffgrade*female) + 
                    factor(background) + #( j)
                    factor(country) + female*factor(country) + #( k)
                    I(lgdppc*female) + #(GDP * Female)
                    fisced + I(fisced*female) + misced + I(misced*female) + dadwork + I(dadwork*female
                    homepos + I(homepos*female), #Home Possessions

                    data = mathData, weights = stweight)
model5$coefficients = model5$coefficients[c(1:6,53:63)]
model5$coefficients
```

```
##         (Intercept)              female     I(ggi * female)                 age
##         279.3700519         -96.3542708         186.4084831           6.5817149
##     I(age * female)            diffgrade  I(lgdppc * female)              fisced
##           1.2321770         -10.6268202          -4.3617829           6.0635105
##  I(fisced * female)              misced  I(misced * female)             dadwork
##          -2.8528687           2.1688422           0.1848889          19.7245562
## I(dadwork * female)             momwork I(momwork * female)               homepos
##          -7.1229655          14.9310291         -11.9700067           9.7680502
## I(homepos * female)
##           8.7850346
```

```
#Model 6: Included School Influence [to model 5]
model6 = lm(pv1math ~ female + I(ggi*female) + age + I(age*female) + diffgrade + I(diffgrade*female) + 
                    factor(background) + #( j)
                    factor(country) + female*factor(country) + #( k)
                    I(lgdppc*female) + #(GDP * Female)
                    fisced + I(fisced*female) + misced + I(misced*female) + dadwork + I(dadwork*female
                    homepos + I(homepos*female) + #Home Possessions
                    pcgirls + I(pcgirls*female) + private + I(private*female) + metropolis + I(metrop

                    data = mathData, weights = stweight)
model6$coefficients = model6$coefficients[c(1:6,52:68)]
model6$coefficients
```

```
##           (Intercept)                 female        I(ggi * female)
##            289.593631              -8.216078             151.954750
##                   age         I(age * female)              diffgrade
##              6.360967              -3.767499             -15.746678
##     I(lgdppc * female)                 fisced     I(fisced * female)
##             -4.599992               5.624462              -3.058672
##                misced      I(misced * female)                dadwork
##              1.932381               1.360008              22.399941
##    I(dadwork * female)                momwork    I(momwork * female)
##            -10.579639              12.906018             -11.028071
##               homepos     I(homepos * female)               pcgirls
##              8.025186              10.432228              -8.074682
##    I(pcgirls * female)                private    I(private * female)
##             39.543384               4.381192               4.451629
##            metropolis I(metropolis * female)
##             18.171734             -16.025768
```

```
stargazer(model1, model2, model3, model4, model5, model6, title = "Table 1 - Gender Equality and the Ma
          covariate.labels= c("Female", "GGI x Female", "Age of Student", "Age x Female", "Diff. Grade"
                              "GDP x Female", "GGI", "GDP",
                              "Dad educ.", "Dad educ. x Female", "Mom educ.", "Mom educ. x Female", "Dad
                              "Home possessions", "Home possessions x Female",
                              "Proportion of Girls at School", "Prop. girls x Female", "Private school"
                              "School is in a metropolis", "School is in a Metro x Female"))
```

```
##
## Table 1 - Gender Equality and the Math Gender Gap
## ==================================================================================================
##
##                               ---------------------------------------------------------------------
##
##                                          (1)                        (2)                        (3)
## --------------------------------------------------------------------------------------------------
## Female                                -164.648*                  -154.427*                  -86.
##                                       (92.312)                   (92.462)                   (96.8
##
## GGI x Female                          109.162***                 154.470***                 147.08
##                                       (34.835)                   (42.369)                   (44.3
##
## Age of Student                          4.230                      4.374                      5.9
##                                        (4.071)                    (4.071)                    (4.2
##
## Age x Female                            5.358                      5.187                      1.4
##                                        (5.648)                    (5.648)                    (5.9
##
## Diff. Grade                           -14.103***                 -14.255***                 -17.14
##                                        (2.539)                    (2.540)                    (2.6
##
## Diff. Grade x Female                                              -4.575                     -5.2
##                                                                   (3.539)                    (3.7
##
## GDP x Female                                                                                 103.66
##                                                                                              (3.3
##
```

```
## GGI                                                                                                         3.67
##                                                                                                             (2.89
##
## GDP
##
##
## Dad educ.
##
##
## Dad educ. x Female
##
##
## Mom educ.
##
##
## Mom educ. x Female
##
##
## Dad work
##
##
## Dad work x Female
##
##
## Mom work
##
##
## Mom work x Female
##
##
## Home possessions
##
##
## Home possessions x Female
##
##
## Proportion of Girls at School
##
##
## Prop. girls x Female
##
##
## Private school
##
##
## Private school x Female
##
##
## School is in a metropolis
##
##
## School is in a Metro x Female         370.279***                  365.991***                  175.86
##                                        (64.966)                    (64.999)                    (69.97
##
```

```
## ------------------------------------------------------------------------------------------------
## Observations                              11,527                      11,527                    11,5
## R2                                         0.349                       0.349                     0.2
## Adjusted R2                                0.346                       0.346                     0.2
## Residual Std. Error              319.348 (df = 11468)        319.313 (df = 11467)        335.130 (df
## F Statistic                    106.109*** (df = 58; 11468) 104.394*** (df = 59; 11467) 160.845*** (df
## ================================================================================================
## Note:
```