

R Camp 2019 HW2

Alex Pugh

8/6/2019

Problem 1

I am interested in the impact of leader transitions on economic conditions, specifically foreign direct investment (FDI). When leaders change, it can create uncertainty surrounding future policies. Businesses in other countries may be less willing to make an investment in a country if there is a high level of uncertainty surrounding the future economic policies and tax policies of a country. Not all leader changes generate the same amount of uncertainty and change. If we operate under the assumption that leaders need support to gain and maintain office and thus choose policies that appeal to their source of support, when a leader transition occurs but both leaders come from the same source of support, we should expect less change in policies and a reduced level of uncertainty. On the otherhand, if a new leader comes from a different source of support, there is more uncertainty as to the policies this new support base prefers. Thus, my question is do leader transitions matter independently of changes in source of support (SOLS Changes)?

We are going to create a dataset of the percent change in FDI for one country over 60 years. Theoretically, the change in FDI can take on a positive or negative value bounded between -100% and +100%, though most values would be near 0. The data generating process is influenced by the economic conditions in the country and the uncertainty about policy and future conditions. I will create a variable for GDP growth rate, labour market efficiency score, leader transitions, and SOLS changes. We will also include an aggregate level variable the 5 year average population in millions.

IV1:GDP Growth Rate

First, we will generate a random variable of GDP growth rate for this country. As with change in FDI, the GDP growth rate can take on values between -100% and +100%, though the mean will be closer to 0 than either extreme. I expect that on average the GDP growth rate is more often positive than negative, so I will set the mean to slightly greater than 0.

```
# clear environment
rm(list=ls())

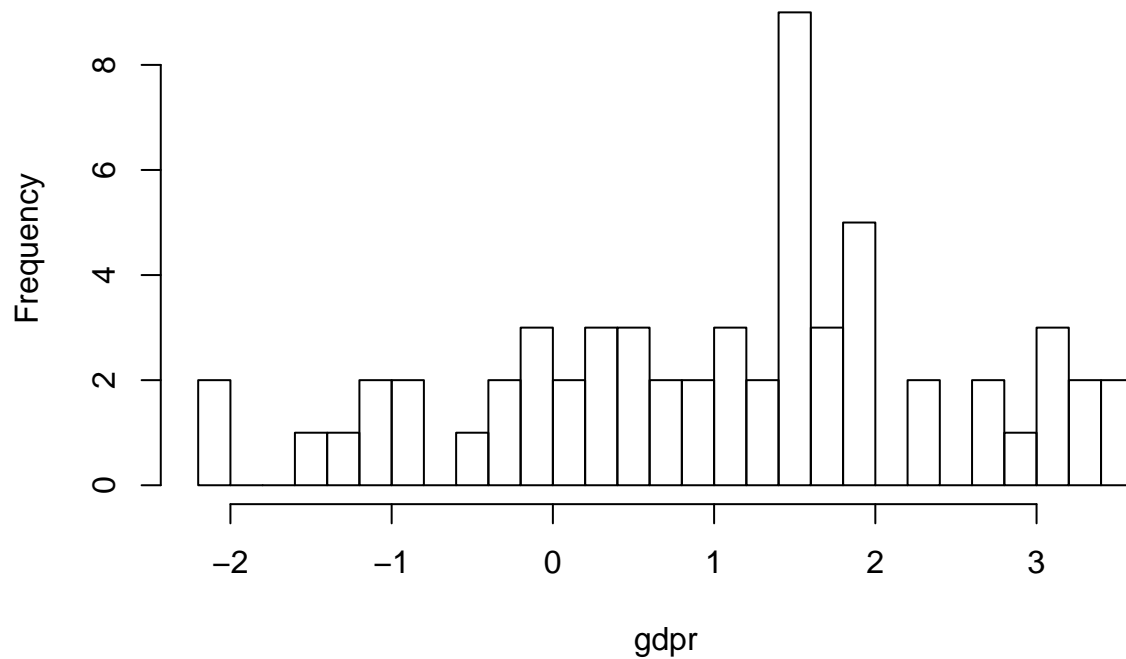
#Create number of disputes for 300
simN <- 60

#generate independent variables
#set seed for random number generator
set.seed(12345)
# create variable of GDP growth rate
gdpr <- rnorm(n=simN, mean=0.8, sd=1.25)
summary(gdpr)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -2.1754  0.1796   1.2514   1.0440  1.8233   3.5460
```

```
hist(gdpr, breaks = 25)
```

Histogram of gdpr



IV2:Labour Market Efficiency Score

The next variable is the labour market efficiency score. This variable can range from 0-100 with higher values indicating greater efficiency. We expect that investors will generally prefer to invest in more developed labour markets. We are going to consider that our country is a developing country, so will not have the highest score.

```
# set random seed
set.seed(1234567)

# create variable with a normal distribution, mean around 18
lme <- rnorm(n=simN, mean=18, sd=3)
summary(lme)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    11.75  16.35   17.55   17.85  19.39   23.93
```

```
cor(gdpr,lme)
```

```
## [1] -0.004715441
```

IV3:Leader Transition

The next variable that we will create is leader transitions. Our country is a mixed system with the effective leader being a prime minister. As such, leader transitions are not fixed at a regular interval. Years where there is no leader transition are going to greater in years than when a leader transition occurs. Also, there can be more than one leader transition in a year, though this is not very common. We will set the maximum number of leader transition to 2.

```

#set seed for random number generator
set.seed(123456)
#create transition variable with most of the observations getting a value of 0.
transition <- rbinom(n=simN, size=2, prob=0.15)
table(transition)

```

```

## transition
## 0 1 2
## 34 24 2

```

IV4:SOLS Change

From the transition variable, we will create is the SOLS change variable. The SOLS change variable is a dummy variable for if a SOLS change occurred in a given year. A SOLS change can only occur in years in which a leadere transition also occurred. Also, if two leader transitions occurred, it is likely that atleast one of them was a SOLS change, so we will code all observations in which there were 2 leader changes as having a SOLS change.

```

set.seed(12345)

#create SOLS change variables based on transition variable
s1 <- rep(0, simN)
s2 <- ifelse(transition == 1 & runif(simN)<0.45, 1, s1) # some transition, SOLS
SOLS <- ifelse(transition > 1, 1, s2) # multiple transitions, SOLS

#Table to check
table(SOLS, transition)

```

```

##      transition
## SOLS 0 1 2
##      0 34 14 0
##      1 0 10 2

```

IV5:Average 5 year Population

The final independent variable we will create is an aggregate variable of the 5 year average population. This variable will be increasing slightly across each 5 year period, though the rate of increase will differ. We expect that the 5 year average population will have a positive relationship with the FDI rate.

```

#set seed
set.seed(12345)

#create averages
num <- simN/5
pop <- rnorm(n=num, mean = 30, sd =3)
pop <- sort(pop)
pop

```

```

## [1] 24.54613 27.24203 28.63951 29.14752 29.17145 29.65126 29.67209
## [8] 31.75659 31.81766 31.89030 32.12840 35.45194

```

```

#create loop to apply to every 5 years
#create new variable
population <- rep(pop, 5)
population <- sort(population)
#Verify that there are 5 of each value
table(population)

```

```

## population
## 24.5461320968888 27.2420339925776 28.6395084796117 29.1475207681699
##           5           5           5           5
## 29.1714476843244 29.651256580944 29.6720900559568 31.7565864535316
##           5           5           5           5
## 31.8176623675212 31.8902956532052 32.1283980525286 35.4519361311127
##           5           5           5           5

```

DV:FDI rate

Now we can create our dependent variable from the independent variables.

```

# create year variable from 1-60
year <- seq(from = 1, to = simN, by = 1)
# create data frame
fdidat <- data.frame(year, gdpr, transition, SOLS, lme, population)

#set seed for random number generator
set.seed(123456)
#create error variable
error <- rnorm(n=simN, mean=0, sd=1)

# set coefficients
b.gdpr <- runif(n=simN, min=0.1, max=0.1)
b.lme <- runif(n=simN, min=0.2, max=0.2)
b.transition <- runif(n=simN, min=-0.4, max=-0.4)
b.sols <- runif(n=simN, min=-0.8, max=-0.8)
b.population <- runif(n=simN, min=0.02, max=0.02)

# create fdir with equation
fdir <- b.gdpr*gdpr + b.lme*lme + b.transition*transition + b.sols*SOLS +b.population*population +error

#ensure within reasonable range
summary(fdir)

```

```

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  1.343   3.205   3.992   3.990   4.588   7.239

```

```

#add to dataframe
fdidat$fdir <- fdir

#create factor variables for transition and SOLS
fdidat$TRANS <- factor(fdidat$transition, levels = c(0,1,2), labels = c("No Transition", "1 Transition", "2 Transition"))
fdidat$sols <- factor(fdidat$SOLS, levels = c(0,1), labels = c("No SOLS", "SOLS"))

```

Descriptive Statistics

Now that we have created our data, we can explore some descriptive statistics

```
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
```

```
#produce table with title, median, and variable labels
```

```
stargazer(fdidat, type="text", title="Summary Statistics", median=TRUE, covariate.labels=c("Year", "GDP (billion USD)", "FDI rate", "5 Year Avg Pop", "SOLS Change", "Labour Market", "Leader Transition", "GDP Growth"))
```

```
##
```

```
## Summary Statistics
```

```
## =====
```

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Year	60	30.500	17.464	1	15.8	30.5	45.2	60
GDP Growth	60	1.044	1.402	-2.175	0.180	1.251	1.823	3.546
Leader Transition	60	0.467	0.566	0	0	0	1	2
SOLS Change	60	0.200	0.403	0	0	0	0	1
Labour Market	60	17.849	2.568	11.753	16.351	17.552	19.385	23.932
5 Year Avg Pop	60	30.093	2.683	24.546	29.021	29.662	31.836	35.452
FDI rate	60	3.990	1.202	1.343	3.205	3.992	4.588	7.239

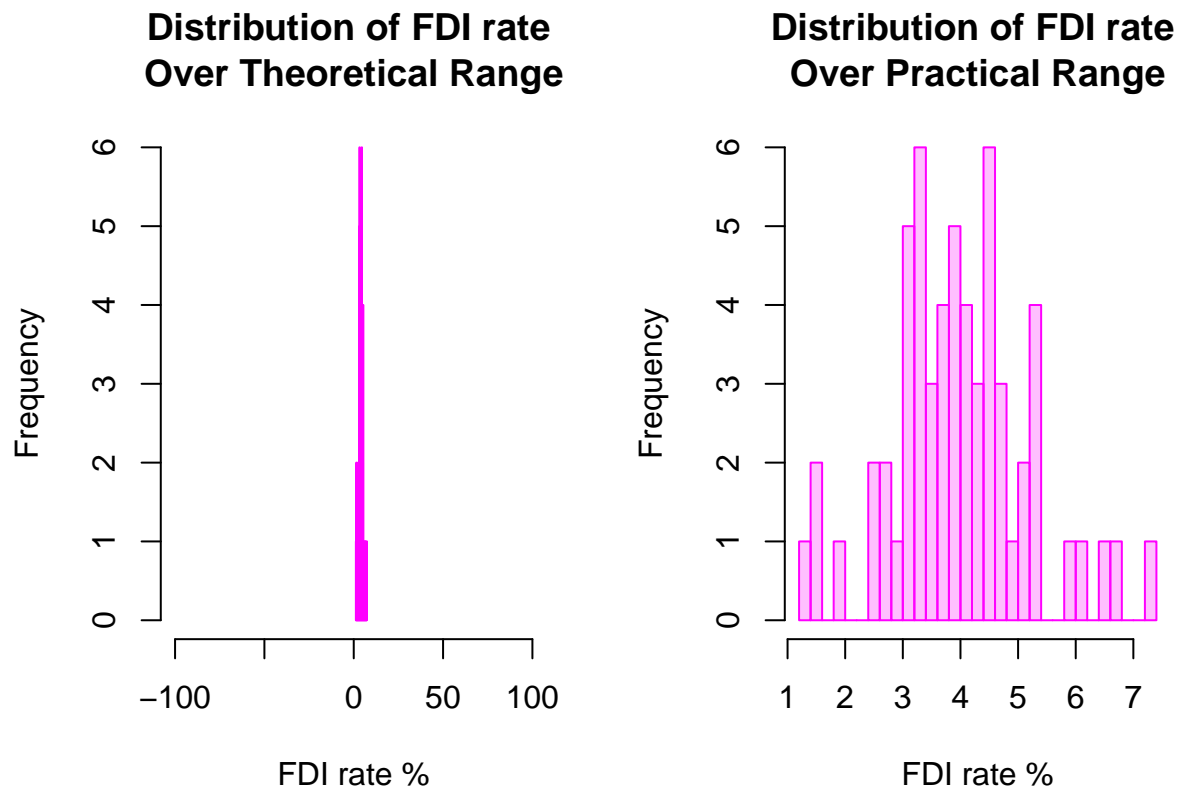
We can plot the distribution of the FDI rate to better understand our dependent variable. Theoretically, the FDI rate as expressed in percents can vary from -100% to +100%, though this is very unlikely. We will still plot over this range as well as the range in our dataset.

```
#create plot with both graphs
```

```
par(mfrow=c(1,2))
```

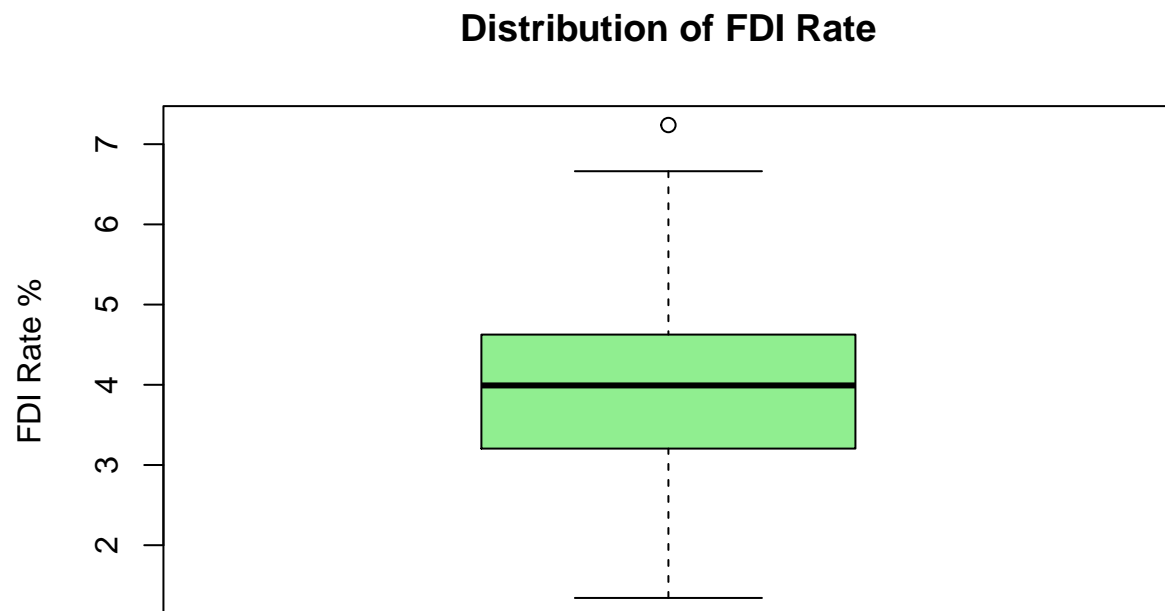
```
hist(fdidat$fdir, breaks=40, col = "#ff00ff40", border = "#ff00ff", xlim=c(-100,100),  
     main="Distribution of FDI rate \nOver Theoretical Range", xlab="FDI rate %")
```

```
hist(fdidat$fdir, breaks=25, col = "#ff00ff40", border = "#ff00ff",  
     main="Distribution of FDI rate \nOver Practical Range", xlab="FDI rate %")
```

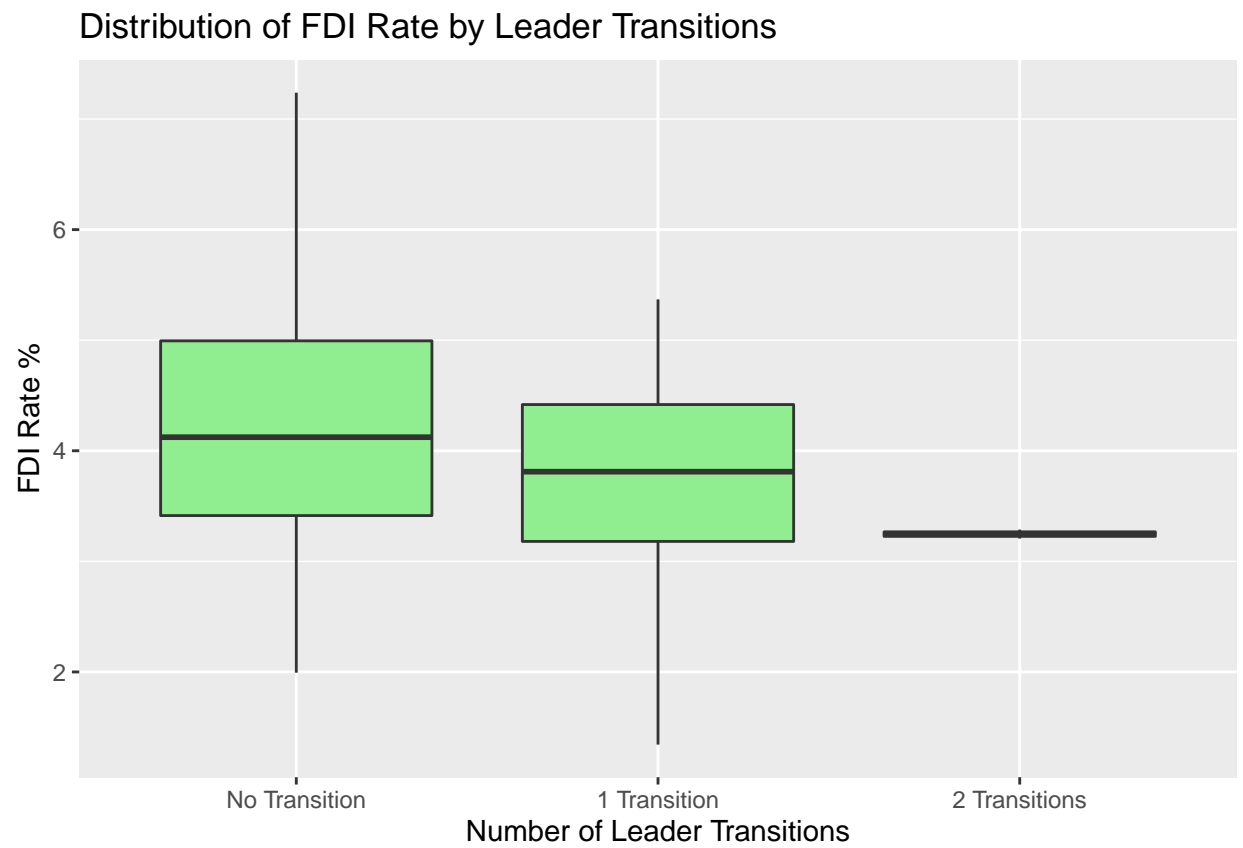


It appears that we may have a few outliers, though it's not completely clear. We can use boxplots to explore further.

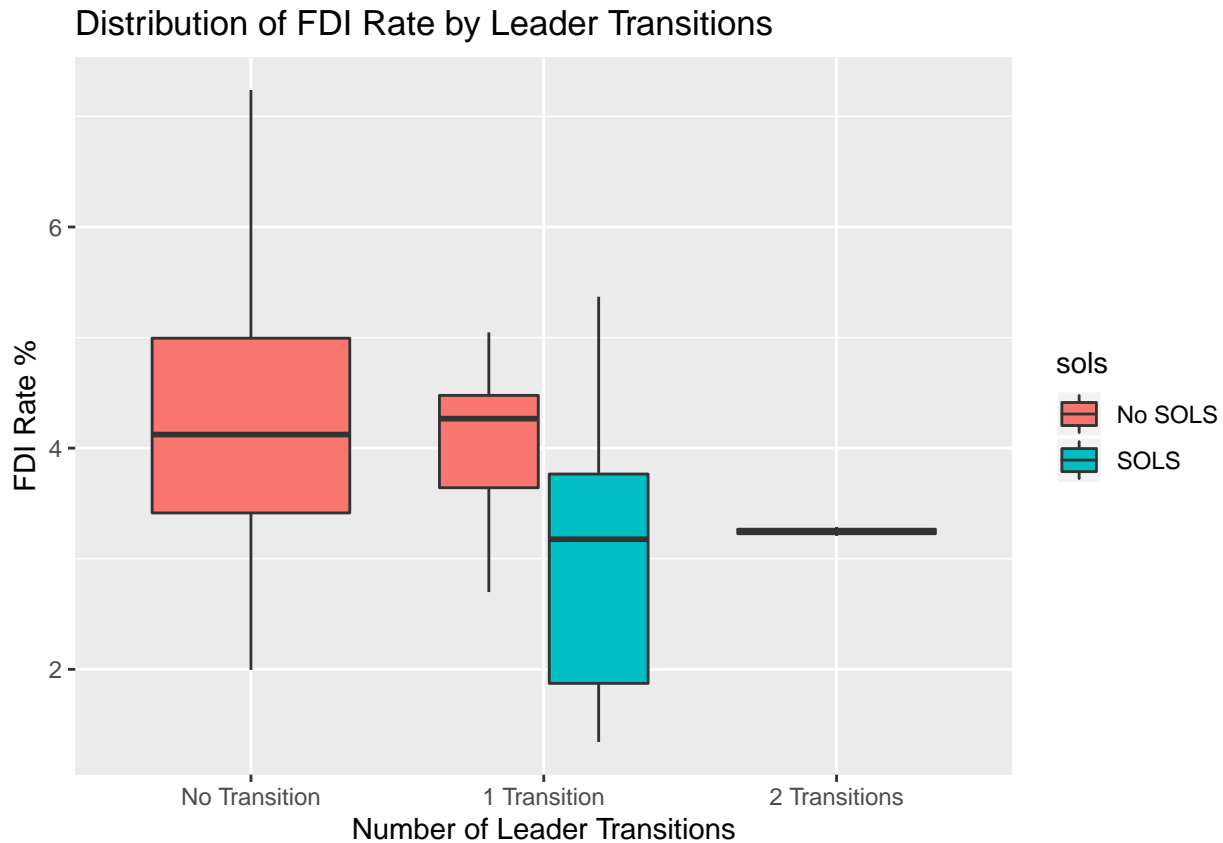
```
library(ggplot2)
library(gridExtra)
#simple boxplot
boxplot(fdidat$fdir, main="Distribution of FDI Rate", ylab="FDI Rate %", col = "light green")
```



```
p1 <- ggplot(fdidat, aes(x=TRANS, y=fdir)) + geom_boxplot(fill="light green") + ggtitle("Distribution of FDI Rate by Leader Transitions")
p1
```



```
p2 <- ggplot(fdidat, aes(x=TRANS, y=fdir, fill=sols)) + geom_boxplot() + ggtitle("Distribution of FDI Rate by Leader Transitions")
p2
```



From these plots, we can see that there may be an outlier. We can also tell that the distribution of the FDI rate does not change much across the number of leader transitions. However, when we include SOLS changes, we can see that when a SOLS change occurs, it appears that the distribution of the FDI rate shifts towards lower values.

We have looked at the distribution of the FDI rate across leader transitions and SOLS changes, but have not addressed the continuous dependent variables, GDP growth rate and labour market efficiency. We can look at the relation between the FDI rate and these through correlation plots and scatter plots.

```
#create correlation matrix and plot
round(cor(fdidat[,2:7]), digit=2) # Rounding numbers
```

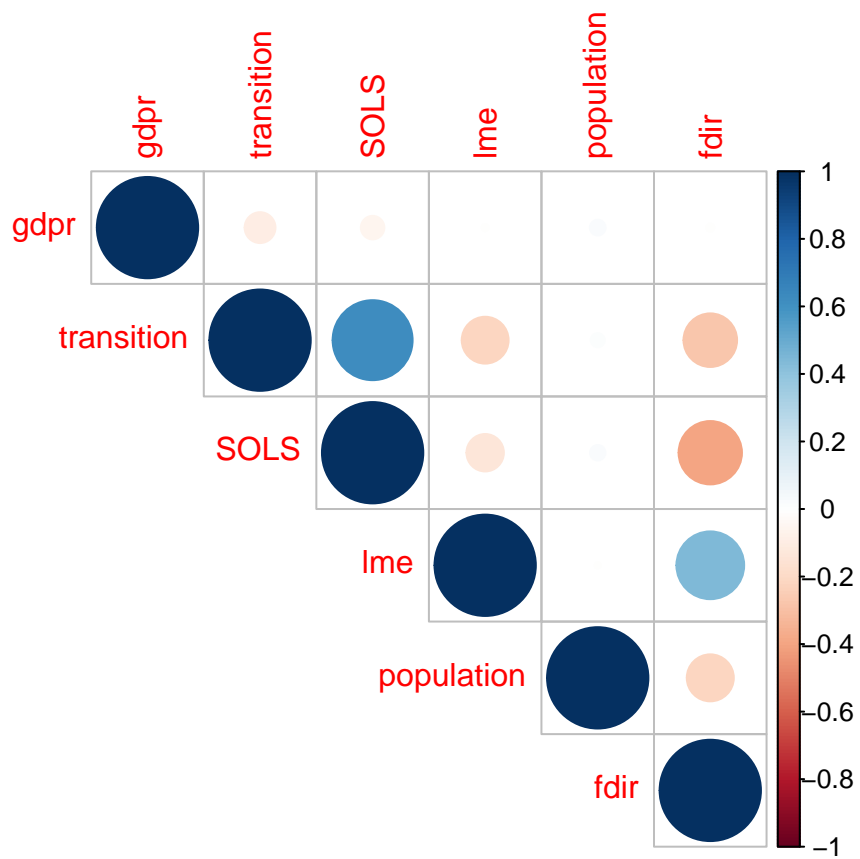
```
##          gdpr transition  SOLS   lme population  fdir
## gdpr      1.00      -0.09 -0.05  0.00         0.02 -0.01
## transition -0.09      1.00  0.62 -0.21         0.02 -0.28
## SOLS       -0.05      0.62  1.00 -0.14         0.02 -0.39
## lme         0.00     -0.21 -0.14  1.00         0.00  0.45
## population  0.02      0.02  0.02  0.00         1.00 -0.22
## fdir       -0.01     -0.28 -0.39  0.45        -0.22  1.00
```

```
library(corrplot)
```

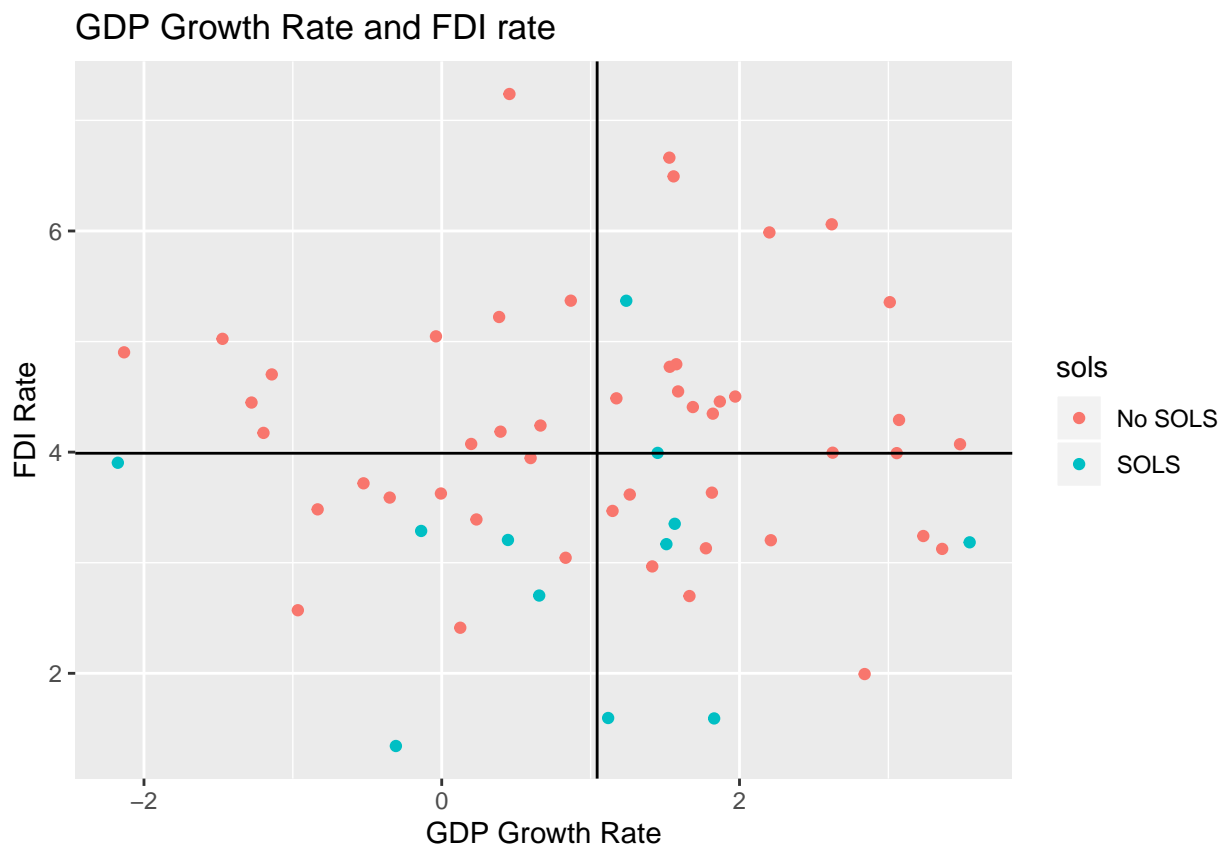
```
## corrplot 0.84 loaded
```



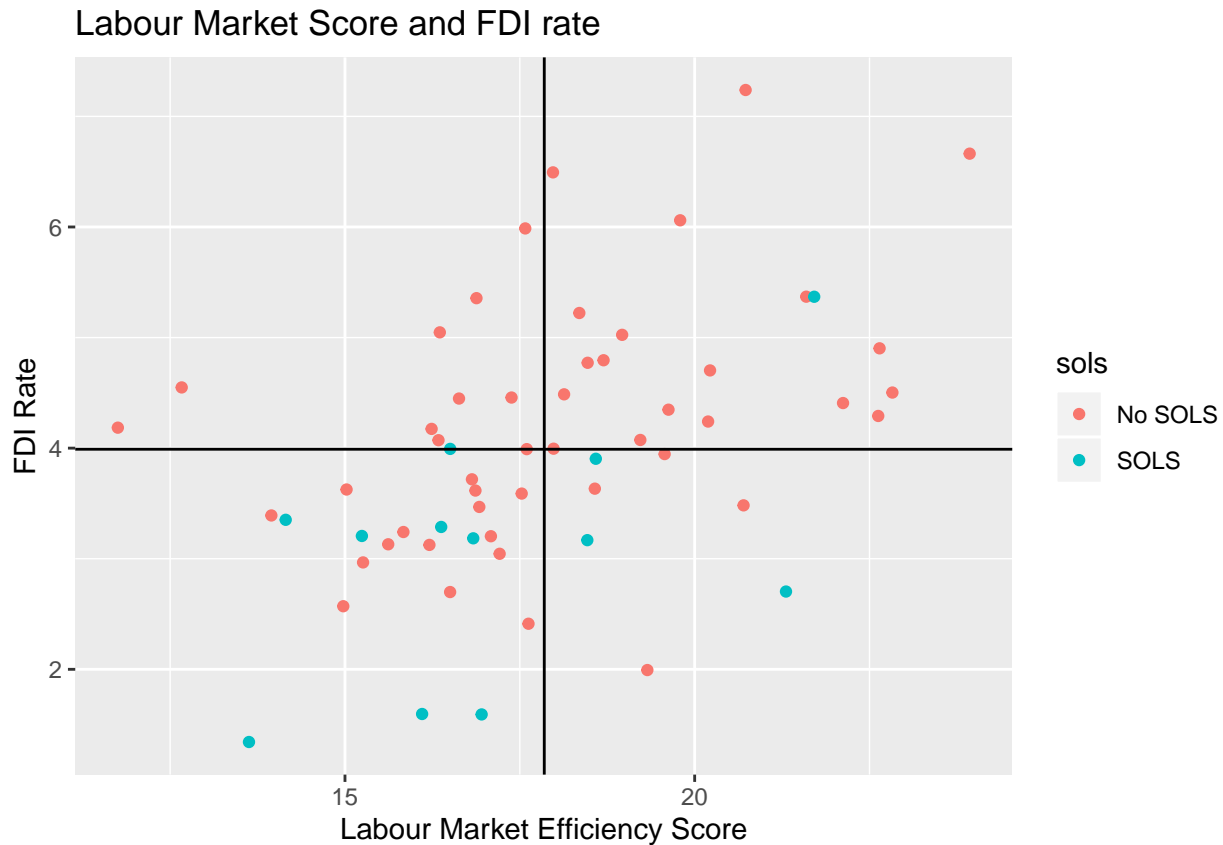
```
co_mat <- cor(fdidat[,2:7]) # Saving correlation matrix
corrplot(co_mat, type="upper")
```



```
meang <- mean(fdidat$gdpr)
meanf <- mean(fdidat$fdir)
meanl <- mean(fdidat$lme)
#create scatterplot with gdpr
ggplot(fdidat, aes(x=gdpr, y=fdir))+geom_point(aes(colour=sols))+ggtitle("GDP Growth Rate and FDI rate")
```



```
#create scatterplot with lme  
ggplot(fdidat, aes(x=lme, y=fdir))+geom_point(aes(colour=sols))+ggtitle("Labour Market Score and FDI r
```



Model

The original coefficients that produced the data were 0.1 for GDP growth rate, 0.2 for the labour market efficiency score, -0.4 for leader transitions, -0.8 for SOLS changes, and 0.02 for population. The model found two of our variables to be statistically significant: labour market efficiency and SOLS changes. Of the coefficient values, the ones that are significant are closest to the true values of the data generating process. The signs for the other variables are all backwards, but the coefficients are near 0 for GDP and population, so that may explain the difficulty in estimating the coefficients. The value for the coefficient for transition is very different from the data generating value. Part of the issue may be the correlation between SOLS changes and leader changes.

```
# OLS (Ordinary Least Squares)
m_ols <- lm(fdir ~ gdpr+lme+transition+SOLS+population, data=fdidat) # Normal-identity model
summary(m_ols)

##
## Call:
## lm(formula = fdir ~ gdpr + lme + transition + SOLS + population,
##     data = fdidat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4725 -0.7136 -0.0835  0.6782  2.2511
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)

```

```
## (Intercept)  3.57782    1.76831    2.023 0.048004 *
## gdpr        -0.01324    0.09425   -0.141 0.888762
## lme         0.18944    0.05243    3.613 0.000665 ***
## transition   0.05650    0.30194    0.187 0.852262
## SOLS        -1.03841    0.41686   -2.491 0.015844 *
## population  -0.09217    0.04904   -1.880 0.065572 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.01 on 54 degrees of freedom
## Multiple R-squared:  0.354, Adjusted R-squared:  0.2942
## F-statistic: 5.919 on 5 and 54 DF,  p-value: 0.0001967
```

```
library(xtable)
xtable(m_ols)
```

```
## % latex table generated in R 3.6.1 by xtable 1.8-4 package
## % Wed Aug 21 13:12:52 2019
## \begin{table}[ht]
## \centering
## \begin{tabular}{rrrrr}
## \hline
## & Estimate & Std. Error & t value & Pr(>|t|) \\
## \hline
## (Intercept) & 3.5778 & 1.7683 & 2.02 & 0.0480 \\
## gdpr & -0.0132 & 0.0942 & -0.14 & 0.8888 \\
## lme & 0.1894 & 0.0524 & 3.61 & 0.0007 \\
## transition & 0.0565 & 0.3019 & 0.19 & 0.8523 \\
## SOLS & -1.0384 & 0.4169 & -2.49 & 0.0158 \\
## population & -0.0922 & 0.0490 & -1.88 & 0.0656 \\
## \hline
## \end{tabular}
## \end{table}
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.5778	1.7683	2.02	0.0480
gdpr	-0.0132	0.0942	-0.14	0.8888
lme	0.1894	0.0524	3.61	0.0007
transition	0.0565	0.3019	0.19	0.8523
SOLS	-1.0384	0.4169	-2.49	0.0158
population	-0.0922	0.0490	-1.88	0.0656

We can plot the predicted FDI rate across labour market efficiency scores, holding other covariates at their median. The plots demonstrates that as scores increase, the predicted FDI rate increases.

```
#create plot of predicted values across labour market scores
#set mean values of covariates
GDPR <- median(fdidat$gdpr)
POP <- median(fdidat$population)
S <- median(fdidat$SOLS)
LT <- median(fdidat$transition)
```

```

#create x axis variable
xrange <- seq(from = 10, to = 25, by = 0.5)
#get coefficients
coefs <- coef(m_ols)
coefs

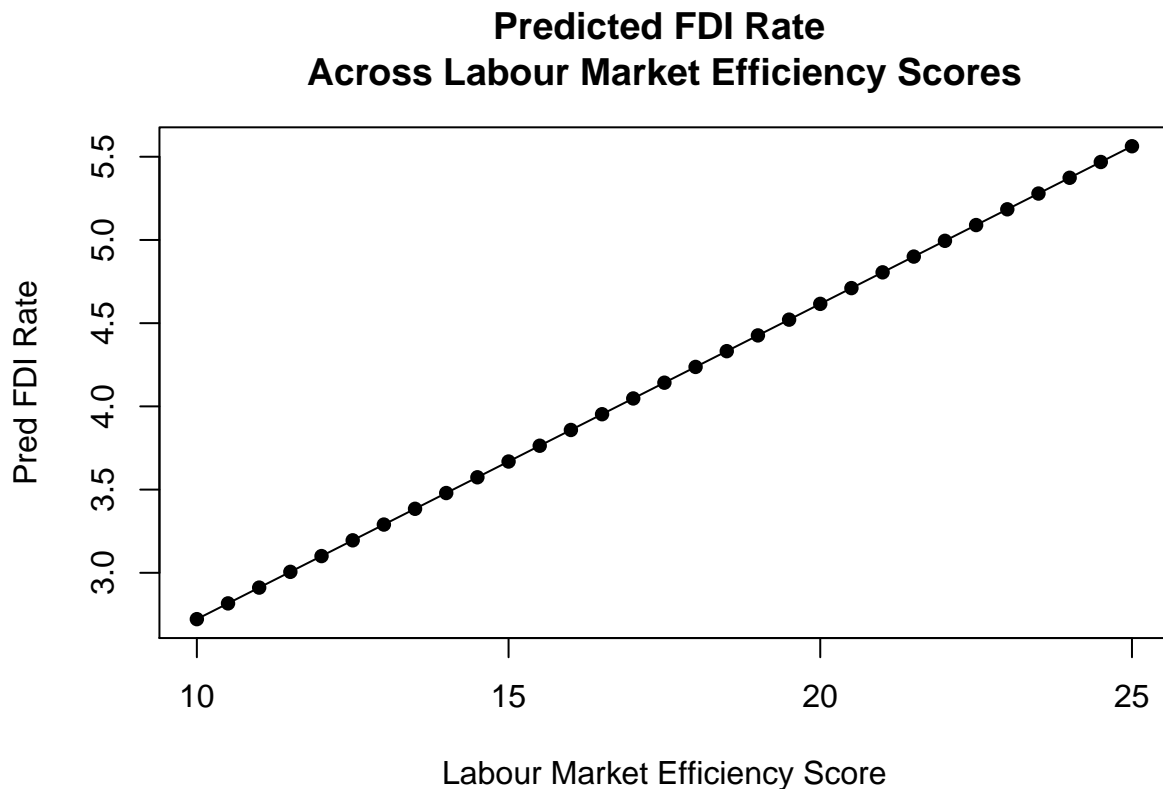
## (Intercept)          gdpr          lme  transition          SOLS  population
## 3.57781579 -0.01324465  0.18943784  0.05650123 -1.03841157 -0.09217441

#create empty vector for predicted values
pred <- rep(NA,length(xrange))

index <- 1
#create linear aggregator
for(i in xrange){
  LA <- coefs[1] + GDPR * coefs[2] + i * coefs[3] + LT * coefs[4] + S * coefs[5] + POP * coefs[6]
  pred[index] <- LA
  index <- index + 1
}

#plot vlaues of linear aggregator across
plot(pred ~ xrange, ylab="Pred FDI Rate", pch=16,
      xlab="Labour Market Efficiency Score")
lines(pred ~ xrange, type="l")
title("Predicted FDI Rate \nAcross Labour Market Efficiency Scores")

```



While we were not able to replicate the full data generating process, we captured the effect of SOLS changes, a key variable of interest, fairly well. Increasing the number of observations may allow us to better replicate the data generating process.

Comments

If I were actually to collect this data, I would have to be aware of time dependencies. I choose to ignore them in this exercise, but the FDI rate in one year may be highly correlated with the FDI rate from the previous year. Also, I might want consider the temporal effect of my independent variables. Does it take a year for the GDP growth rate to have an effect? What is the temporal impact of a SOLS change; is it only for the year of the SOLS change or does it last a little longer? I would also need to consider what the labour market score is capturing and how that is related to GDP growth rate or even whether there is some reverse causality at play.

Problem 2

When states have a dispute, there are multiple outcomes at which they can arrive. Bargaining failure over a dispute may merely result in a stalemate, or it may bring the states into conflict. Since the World War Era, the international community has fostered the formalization of a third option: arbitration or adjudication. My question is when do states choose to go to war as opposed to seeking peaceful settlement through arbitration or adjudication. For this purposes of this exercise, we will ignore the possibility of stalemates. We will also assume that both states have the ability to unilaterally appeal to the court, so we can look at this decision from the perspective of one state.

The data generating process for our dependent variable of Court/War is determined by an underlying utility function. When the utility of going to the court is greater than 0, we observe that state taking the dispute to court. I expect that in this world with only two options of dispute settlement, war or adjudication, we would see more adjudication because it is generally less costly.

The utility of going to the court is likely determined by the likelihood of a favorable ruling, whether a state is a democracy, and the costs of the court. If a state expects a favorable ruling, they are more likely to take a dispute before the court. Democracies are also potentially more likely to take a case to court because of their higher domestic costs for war and their norms surround rule of law. However, there are costs of going to the court in terms of financial costs and opportunity costs of a lack of a solution as a case drags on. If the costs are too high, states may be less likely to take a case before the court. We will first randomly generate values for each of these variables and then create our dependent variable.

IV1: Favorable Ruling

First, we will generate a variable that takes a 1 if a state believes that if it took the dispute to court, the ruling would be favorable and a 0 if it thinks that it would be unfavorable. I expect that states are generally slightly distrustful of the court and expect unfavorable rulings more often than favorable rulings.

```
# clear environment
rm(list=ls())

#set seed for random number generator
set.seed(123456)

#Create number of disputes for 300
simN <- 300

#generate independent variables

# create variable if likely to be favorable vote (0,1)
favorable <- rbinom(n=simN, size=1, prob=0.45)
table(favorable)
```

```
## favorable
##    0    1
## 168 132
```

IV2: Democracy

The second independent variable is a dummy variable for democracy. Of the states that have disputes, I expect that more will be non-democracies rather than democracies.

```
#set seed for random number generator
set.seed(123456)

#generate democracy
demo <- rbinom(n=simN, size=1, prob=0.40) # 40% = demo
# ensure that there are no values greater than 100 or less than 0
table(demo)
```

```
## demo
##    0    1
## 184 116
```

IV3: Court Costs

The next random independent variables we will create are the costs of court. We will depict the cost variable as a categorical variable with 3 categories: Low Cost, Average Cost, High Cost. I expect that for most states, the costs of going to the court are average.

```
#set seed for random number generator
set.seed(123456)

court <- rbinom(n=simN, size=2, prob=0.50)

summary(court)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0000  0.0000  1.0000  0.9967  2.0000  2.0000
```

```
table(court)
```

```
## court
##    0    1    2
##  81 139  80
```

IV4: World Region

Finally, we create a world region variable. Disputes typically take place between neighbors or close trading partners. Within a region, there may be some norms about dispute settlement or shared values regarding avoiding conflict, but these values likely vary across regions. I have created a variable that divides the world into 6 regions. I have ranked the regions on a scale from the least prone to peaceful dispute resolution to those most prone, that way I can include the variable in the model as having a linear effect rather than estimate a coefficient for each region. The order, which does not reflect reality, is North America, Africa, Asia, Europe, South America, Pacific.

```

#set seed for random number generator
set.seed(123456)

region <- rbinom(n=simN, size=5, prob=0.5)

table(region)

## region
##  0  1  2  3  4  5
## 10 47 92 93 53  5

```

DV:Dispute Outcome

Now that we have all of our independent variables, we will create a utility function for the court and then code our dependent variable as observations with a utility greater than 0. We expect that the favorability of a ruling and probability of compliance will have a positive coefficient while the costs of the court will have a negative coefficient.

```

#set seed for random number generator
set.seed(123456)

state <- seq(from=1, to = simN, by=1)

#create data frame of variables
sim.cw <- data.frame(state, favorable, court, demo, region)

#view first lines
head(sim.cw, n=2)

```

```

##      state favorable court demo region
## 1         1         1     2     1     3
## 2         2         1     2     1     3

```

```

#set seed for random number generator
set.seed(123456)
#error term for both
errorc<- rnorm(n=simN, mean=0, sd=1)
summary(errorc)

```

```

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -2.74887 -0.62808  0.04544  0.02738  0.72913  2.58773

```

```

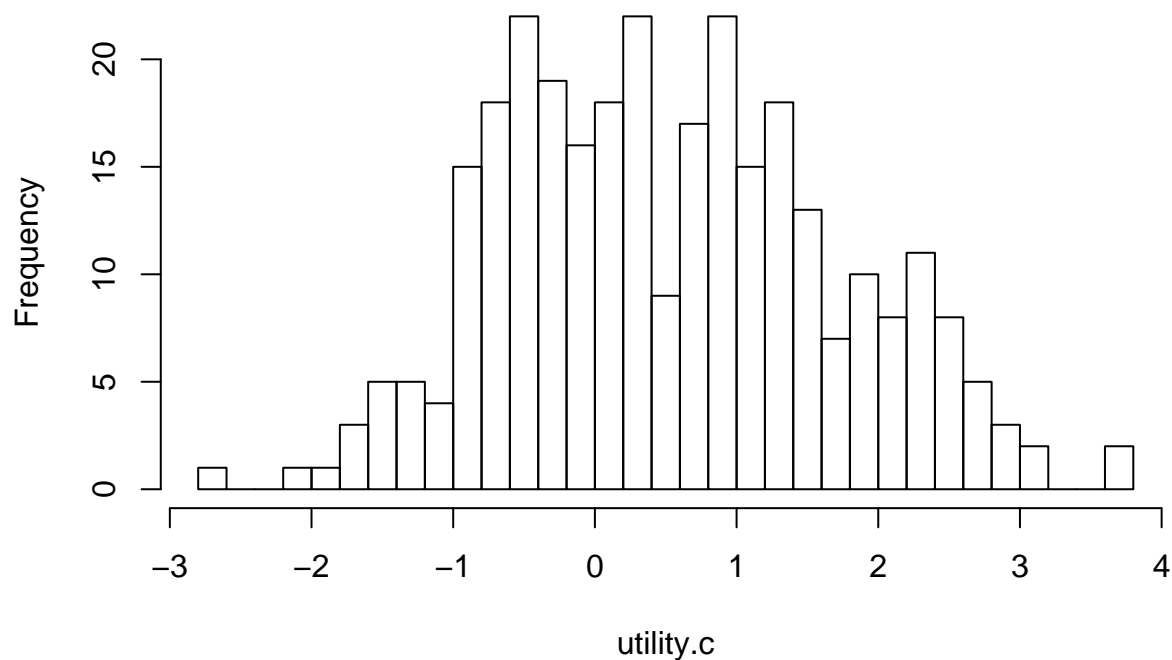
#create coefficients for each
b.ruling <- runif(n=simN, min=0.6, max=0.6) # COEF OF ruling: START FROM A SIMPLE MODEL
b.court  <- runif(n=simN, min=-0.2, max=-0.2) # COEF OF costs of the court: START FROM A SIMPLE MODEL
b.demo <- runif(n=simN, min=0.8, max=0.8) # COEF OF Democracy: START FROM A SIMPLE MODEL
b.region <- runif(n=simN, min=0.05, max=0.05) # COEF OF region: START FROM A SIMPLE MODEL

#Create utility functions
utility.c <- b.ruling*favorable + b.court*court + b.demo*demo+ b.region*region + errorc # UTILITY FUNCTION

hist(utility.c, breaks = 40, main="Distribution of Court Utility Function")

```


Distribution of Court Utility Function



```
# create binary dependent variable, coding 1 if utility function greater than 0
lawsuit <- ifelse(utility.c > 0, 1, 0)

sim.cw$lawsuit <- lawsuit

# create factor variables for ease in graphing
sim.cw$outcome <- factor(sim.cw$lawsuit, levels= c(0,1), labels = c("War", "Court"))
sim.cw$costscourt <- factor(sim.cw$court, levels= c(0,1,2), labels = c("Low Cost", "Average Cost", "High Cost"))
sim.cw$democracy <- factor(sim.cw$demo, levels= c(0,1), labels = c("Non-Democracy", "Democracy"))
sim.cw$fav <- factor(sim.cw$favorable, levels= c(0,1), labels = c("Unfavorable", "Favorable"))
sim.cw$Region <- factor(sim.cw$region, levels= c(0,1,2,3,4,5),
  labels = c("North America", "Africa", "Asia", "Europe", "South America", "Pacific"))
```

Descriptive Statistics

Now that we have our simulated data, we will inspect the data.

```
library(stargazer)

#produce table with title, median, and variable labels
stargazer(sim.cw, type="text", title="Summary Statistics", median=TRUE, covariate.labels=c("State", "Prison"))

##
## Summary Statistics
## =====
## Statistic      N    Mean   St. Dev. Min Pctl(25) Median Pctl(75) Max
## -----
## State          300 150.500   86.747    1    75.8    150.5    225.2    300
```

```
## Prob. Ruling 300 0.440 0.497 0 0 0 1 1
## Court Costs 300 0.997 0.734 0 0 1 2 2
## Democracy 300 0.387 0.488 0 0 0 1 1
## Region 300 2.490 1.105 0 2 3 3 5
## Lawsuit 300 0.633 0.483 0 0 1 1 1
## -----
```

There are more instances of a state choosing adjudication over war in the dataset, but we can't tell the distribution from these statistics. The table and plot below demonstrate that a vast majority of our simulated disputant states choose the court over war. We can also see how the dispute outcomes are distributed across the different cost categories.

```
#make frequency table for Outcome variable an object called freq.out
freq.out <- table(sim.cw$outcome)

#make proportion table for outcomea variable an object called prop.out
prop.out <- prop.table(freq.out)

#use cbind to combine objects to display table of frequency and proportion of outcome variable
cbind(freq.out, prop.out)
```

```
##      freq.out  prop.out
## War         110 0.366667
## Court        190 0.633333
```

```
library(ggplot2)

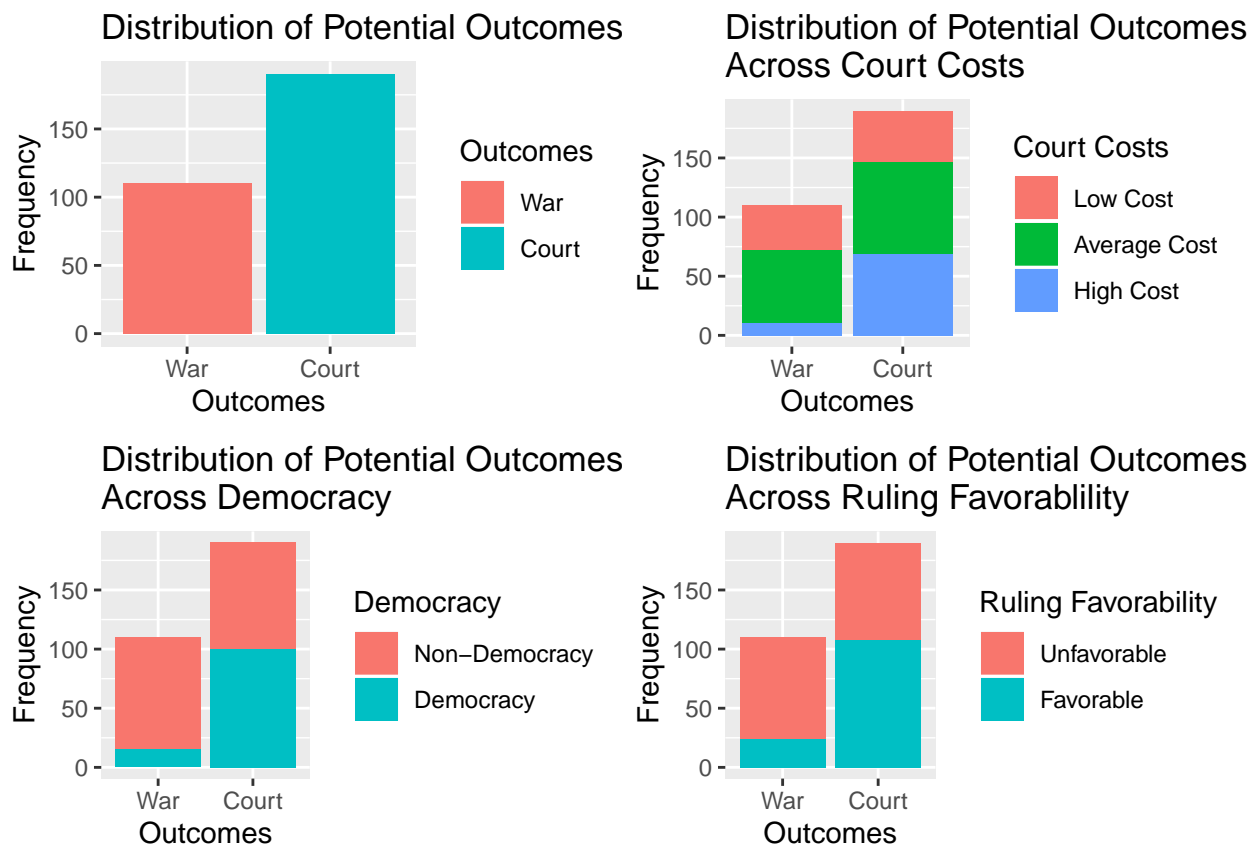
#create barplot
g1 <- ggplot(sim.cw, aes(x=outcome, fill = outcome)) + geom_bar() + ggtitle("Distribution of Potential Outcomes") + xlab("Outcomes") + ylab("Frequency") + labs(fill = "Outcomes")

#create barplot
g2 <- ggplot(sim.cw, aes(x=outcome, fill = costscourt)) + geom_bar() + ggtitle("Distribution of Potential Outcomes by Court Costs") + xlab("Outcomes") + ylab("Frequency") + labs(fill = "Court Costs")

g3 <- ggplot(sim.cw, aes(x=outcome, fill = democracy)) + geom_bar() + ggtitle("Distribution of Potential Outcomes by Democracy") + xlab("Outcomes") + ylab("Frequency") + labs(fill = "Democracy")

g4 <- ggplot(sim.cw, aes(x=outcome, fill = fav)) + geom_bar() + ggtitle("Distribution of Potential Outcomes by Favorability") + xlab("Outcomes") + ylab("Frequency") + labs(fill = "Ruling Favorability")

library(gridExtra)
grid.arrange(g1,g2,g3,g4, ncol =2)
```



Model

Next, we will run a bernoulli logistic model. The coefficients that produced the data were 0.6 for favorable ruling, 0.8 for democracy, -0.2 for court costs, and 0.05 for region. The model produced estimates that were similar to the data generating process, though all greater in value. All of our coefficients are statistically significant at conventional levels.

```
# Bernoulli-logistic model
m_glm_logit <- glm(lawsuit ~ favorable + demo + court + region, family="binomial", data=sim.cw)
summary(m_glm_logit)
```

```
##
## Call:
## glm(formula = lawsuit ~ favorable + demo + court + region, family = "binomial",
##      data = sim.cw)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1141  -1.0811   0.5496   1.1449   1.3002
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.28423    0.37914  -0.750  0.453458
## favorable     -0.07698    0.58050  -0.133  0.894508
## demo          2.06804    0.60604   3.412  0.000644 ***
## court        -0.56211    0.37833  -1.486  0.137342
```

```
## region      0.30777    0.26156    1.177 0.239332
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 394.29  on 299  degrees of freedom
## Residual deviance: 345.79  on 295  degrees of freedom
## AIC: 355.79
##
## Number of Fisher Scoring iterations: 4
```

```
library(xtable)
xtable(m_glm_logit)
```

```
## % latex table generated in R 3.6.1 by xtable 1.8-4 package
## % Wed Aug 21 13:12:54 2019
## \begin{table}[ht]
## \centering
## \begin{tabular}{rrrrr}
## \hline
## & Estimate & Std. Error & z value & Pr(>|z|) \\
## \hline
## (Intercept) & -0.2842 & 0.3791 & -0.75 & 0.4535 \\
## favorable & -0.0770 & 0.5805 & -0.13 & 0.8945 \\
## demo & 2.0680 & 0.6060 & 3.41 & 0.0006 \\
## court & -0.5621 & 0.3783 & -1.49 & 0.1373 \\
## region & 0.3078 & 0.2616 & 1.18 & 0.2393 \\
## \hline
## \end{tabular}
## \end{table}
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.2842	0.3791	-0.75	0.4535
favorable	-0.0770	0.5805	-0.13	0.8945
demo	2.0680	0.6060	3.41	0.0006
court	-0.5621	0.3783	-1.49	0.1373
region	0.3078	0.2616	1.18	0.2393

We set the coefficients to be 0.5 for compliance, 1.1 for favorable ruling, and -0.3 for the costs of the court. The reported coefficients from our model were similar to the coefficients in the data generating process. However, only the costs of court was significant at conventional levels. The high level of correlation between a favorable ruling and compliance may be at fault, making it harder to reject the null hypothesis.

We can plot the predicted probability of going to court across the levels of court costs, holding other covariates at their median. The plots demonstrates that as the costs of the court increase, the predicted probability of going to the court decreases.

```
#create plot of predicted values across Levels of Court Cost
#set median values of covariates
FAV <- median(sim.cw$favorable)
DEM <- median(sim.cw$demo)
REG <- median(sim.cw$region)
```

```

#create x axis variable
xrange <- seq(from = 0, to = 2, by = 1)
#get coefficients
coefs <- coef(m_glm_logit)
coefs

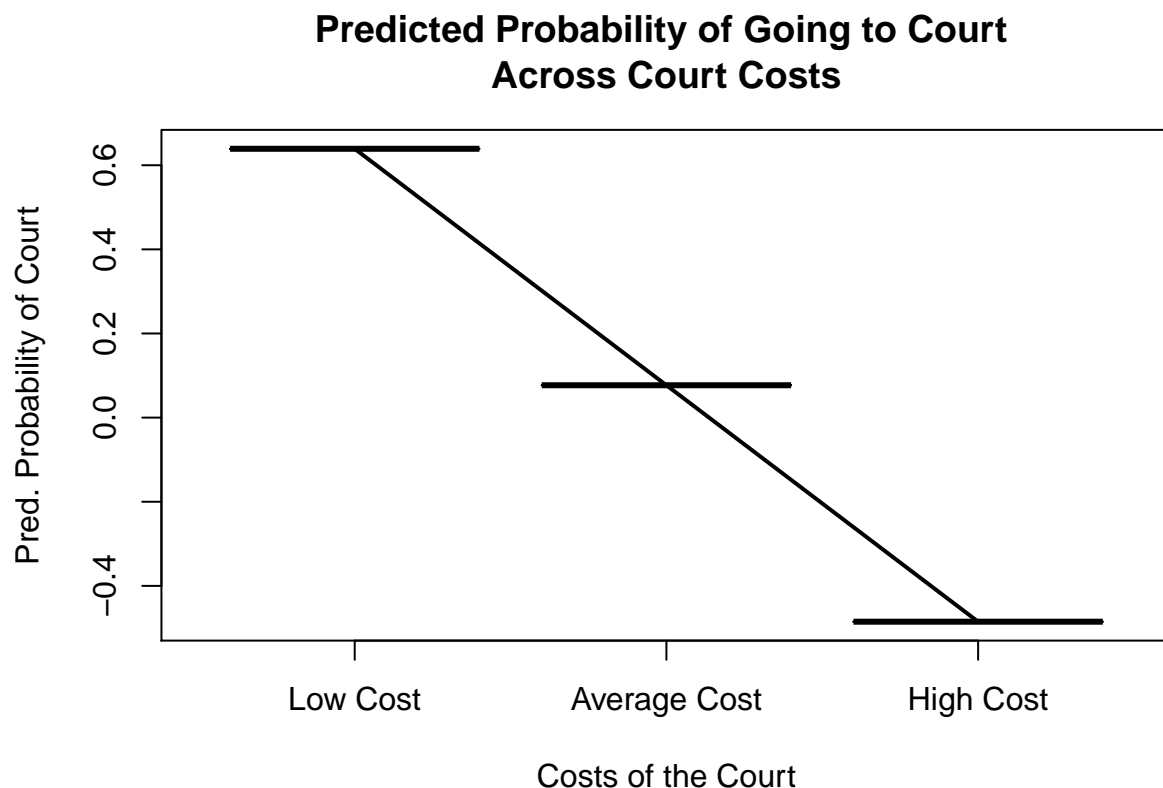
## (Intercept)    favorable        demo        court        region
## -0.28422574 -0.07697606  2.06803636 -0.56211076  0.30777085

#create empty vector for predicted values
pred <- rep(NA,length(xrange))

index <- 1
#create linear aggregator
for(i in xrange){
  LA <- coefs[1] + FAV * coefs[2] + DEM * coefs[3] + i * coefs[4] + REG * coefs[5]
  pred[index] <- LA
  index <- index + 1
}

xrange <- factor(xrange, levels= c(0,1,2), labels = c( "Low Cost","Average Cost", "High Cost"))
#plot vlaues of linear aggregator across
plot(pred ~ xrange, ylab="Pred. Probability of Court", pch=16,
      xlab="Costs of the Court")
lines(pred ~ xrange, type="l", lwd = 2)
title("Predicted Probability of Going to Court \nAcross Court Costs")

```



Comments

If I were to actually collect this data, I would run into quite a few difficulties. First, I would have to define what constitutes a dispute. The ICOW issue claims dataset, once it has been expended, would be helpful in doing this, though it still would not exactly capture my conceptualization of a dispute. Another issue I would have to determine is what constitutes an act of war and what constitutes a move towards adjudication/arbitration. I could use the occurrence of a MID as an act of force, but identifying a move towards adjudication/arbitration would be more difficult since it is typically contingent on the consent of the other state.

Regarding the dependent variables, I would have to find a way to measure the expected favorability of a court ruling. I am aware of measures of strength of legal claims in disputes, so could explore a specification similar to that one. Furthermore, I would need to specify a means of measuring the costs of the court. I would have to explore more the variation in the financial cost as well as determine a way to measure the opportunity cost. Adjudication is not a fast process, so states would have to consider the costs of not reaching a solution.