Determinants of Maritime Piracy

Bj?ern Boening, Cody Koebnick, Laurence Hendry

# Introduction

Maritime piracy has been in the media, especially with attention directed to the Horn of Africa. ATALANTA, the European Union's naval mission, patrolled along the coast of Somalia to combat maritime piracy and protect trading vessels against attacks. As it seems, fading media attention does not presuppose a reduced danger of maritime piracy globally.

Global shipping routes are highly important for trade. Piracy attacks are a potential threat for crew and cargo on the ship. The cost intensive deployment of international naval forces in Somalia shows how serious countries take the threat whose trade is affected. Interestingly not all piracy attacks are successful, and the ratio varies from country to country and over time. So what drives piracy attacks, why and when are they successful?

# Research Design

The idea for this paper is to research a model which can explain piracy under which circumstances attacks are successful.

Does the number of attacks decrease the likelihood of attacks being successful?

The dependent variable is the success rate of piracy attacks, calculated by the number of successful attacks divided by the total number of attacks. We expect that mainly the total number of attacks has an impact on this ratio. The fact that the dependent variable actually consists of our key independent variable is dangerous. However, we think that there must be a visible learning effect, either from law enforcement bodies, the shipping crew, or the pirates. So far this was the only feasible way we could have a look at this effect.

Furthermore, additional exogenous variables will be included that in theory should have an impact on the inspected success rate. GDP (per capita per year) as a mirror for the economic incentives to conduct piracy is expected to influence the success rate over time. Likewise, a country's ratio of coast line to its land area should be a good demarker, whether piracy attacks happen more often.

1. After a certain amount of piracy attacks, the success rate has a peek and will decrease.
2. The higher the GDP, the less incentive there is to conduct an attack and only less skilled pirates will make attempts, hence the success rate decreases.
3. The higher the coast-land ratio is, the more people decide to conduct attacks and skilled pirates emerge, hence the success rate increases.

# Data

The International Maritime Bureau (IMB) collects since 1992 all reported piracy attack globally. Since then it publishes annually an overview of all attacks that happened in a year. These annual reports provided by the IMB contain detailed information about every incident, which allows for further analysis of distinct types of piracy attacks, for instance successful attacks v. attempted attacks.

The annual reports were scraped with text analysis tools. Our team received a "ready to use" dataset from a research project from the university of Tennessee, including all global piracy attacks from 1994 to 2014.

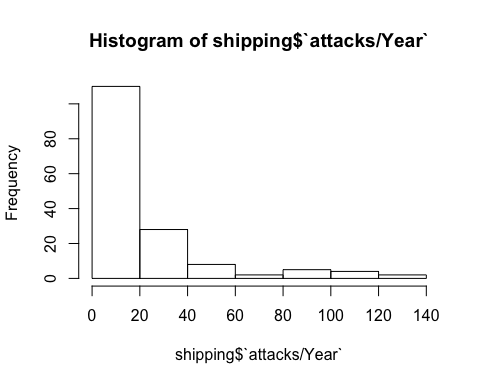
# Additional Data Gathering: Country Coast-Area Ratios

The original dataset contains the attacks that were reported by the victims of piracy. As an additional variable, relevant to our field of piracy investigation and patrols, we were intrigued by the relative effect that a longer coast length of a country has on the level of attacks that country suffers. To address this question we parsed a table titled "List of countries by length of coastline" from a Wikipedia page that had, in turn, used information from the CIA World Factbook. We then merged this coastline data with our existing dataframe using a 'right outer join'.

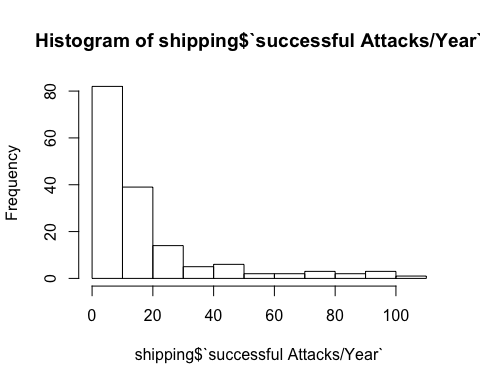
Of critical interest to us were the respective 'Coast/Area' ratios (measured in km of coast length to km of square land) that serves as an insightful control for our country dependent variable.

# Visualizing and examing the data

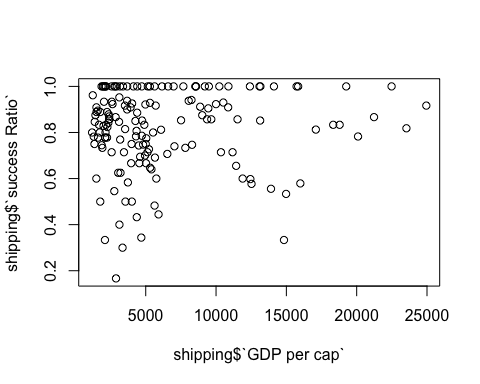
Below is a histogram showing the frequency of attacks sorted by number of attacks per year. In other words, we can clearly see that the vast majority of years had 1-20 attacks reported. The frequency then goes gradually within the 20-40 range and then again in the 40-60 range. We see a slight surge in number of attacks per year within the 80-100 range, this is due to highly pirated coastal areas such as Bangladesh and the philippines.



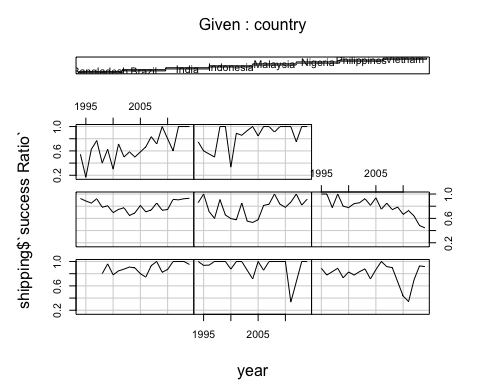
Similarily, the sucess of attacks is distributed in a very similar fashion to the number of attacks. However, our study seeks to see if there is indeed a statistically significant relationship between the number of attacks and attack success.



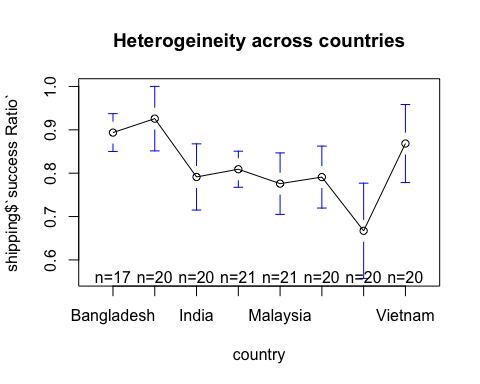
Below, a plot displays the relationship between attack success and the GDP per capita of the closest coastal state.



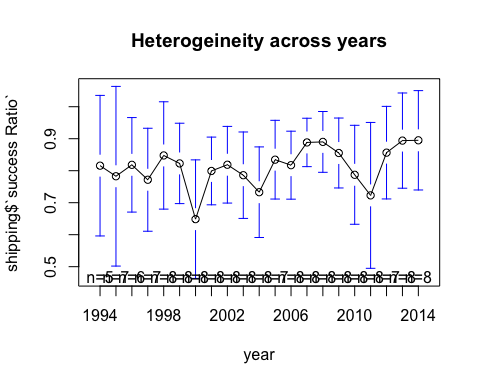
Here we have seperate coplots depicting the success ratio for attacks by year, sorted by the closest coastal state.



When we examine the heterogeneity across countries it appears that a pirates change of a successful attack is dependent on the closest coastal state. For instance, the changes of a successful attack in the Philippines is drastically lower than the odds of a successful attack in Bangladesh. This also holds true when considering confidence intervals which are also depicted.



When we examine the heterogeneity across years the means seem to stay within the .65 - .85 range. However, due to extremely large confidence intervals, time does not appear to be statistically significant.

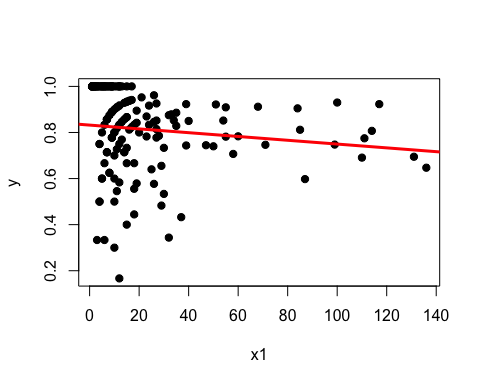


Below is our first OLS regrssion. Although OLS regrssion does not consider heterogeneity across groups or time, an OLS regrssion can still prove useful for gathering initial insight into the relationship of our variables.

For instance, here we see that attacks per year does not a statistically significant effect with a p-value of >.1. However, an interesting point worth noting would be that negative coefficient of attacks per year.

# OLS Regression  
# Regular OLS regression does not consider heterogeneity across groups or time.  
# In this simple model, the number of attacks has a slightly negative relationship with attack success, however it is not stat. sig.   
ols <-lm(shipping$`success Ratio` ~ shipping$`attacks/Year`, data=shipping)  
summary(ols)

##   
## Call:  
## lm(formula = shipping$`success Ratio` ~ shipping$`attacks/Year`,   
## data = shipping)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.65569 -0.07584 0.03410 0.15965 0.18735   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.8322623 0.0180135 46.202 <2e-16 \*\*\*  
## shipping$`attacks/Year` -0.0008251 0.0005090 -1.621 0.107   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1742 on 157 degrees of freedom  
## Multiple R-squared: 0.01646, Adjusted R-squared: 0.0102   
## F-statistic: 2.628 on 1 and 157 DF, p-value: 0.107

The below plot shows that after attacks in a certain country reach a threshold, approximately 40, their attack success ratio is steadily above .6. 

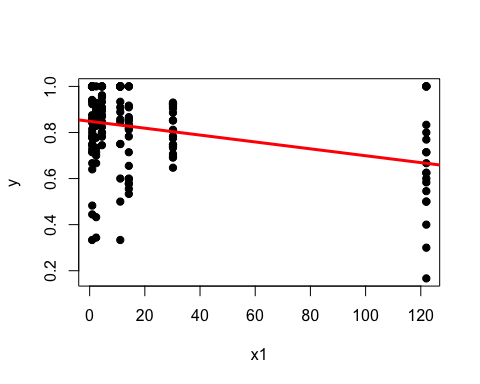
In our model's first fixed effects regression, attacks per year becomes statistically significant and has a small negative coefficient of -.002. It is important to note that the attacks per year became significant only in the fixed effects model, as opposed to the OLS.

## Oneway (individual) effect Within Model  
##   
## Call:  
## plm(formula = shipping$`success Ratio` ~ shipping$`attacks/Year`,   
## data = shipping, model = "within", index = c("country", "year"))  
##   
## Unbalanced Panel: n=8, T=17-21, N=159  
##   
## Residuals :  
## Min. 1st Qu. Median 3rd Qu. Max.   
## -0.6010 -0.0710 0.0331 0.0853 0.3280   
##   
## Coefficients :  
## Estimate Std. Error t-value Pr(>|t|)   
## shipping$`attacks/Year` -0.00231940 0.00080135 -2.8944 0.004366 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Total Sum of Squares: 3.9439  
## Residual Sum of Squares: 3.7353  
## R-Squared : 0.052895   
## Adj. R-Squared : 0.049901   
## F-statistic: 8.37737 on 1 and 150 DF, p-value: 0.0043664

However, once we add a variable controlling for the coast ratio of the closest coastal state, attacks per year once again becomes statistically insignificant.

##   
## Call:  
## lm(formula = shipping$`success Ratio` ~ shipping$`attacks/Year` +   
## shipping$`coast/Area ratio (m/km2)`, data = shipping)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.53257 -0.07033 0.02719 0.13278 0.32096   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 0.8699638 0.0189758 45.846 < 2e-16  
## shipping$`attacks/Year` -0.0009046 0.0004810 -1.881 0.0619  
## shipping$`coast/Area ratio (m/km2)` -0.0015205 0.0003393 -4.482 1.43e-05  
##   
## (Intercept) \*\*\*  
## shipping$`attacks/Year` .   
## shipping$`coast/Area ratio (m/km2)` \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1645 on 156 degrees of freedom  
## Multiple R-squared: 0.1287, Adjusted R-squared: 0.1175   
## F-statistic: 11.52 on 2 and 156 DF, p-value: 2.162e-05

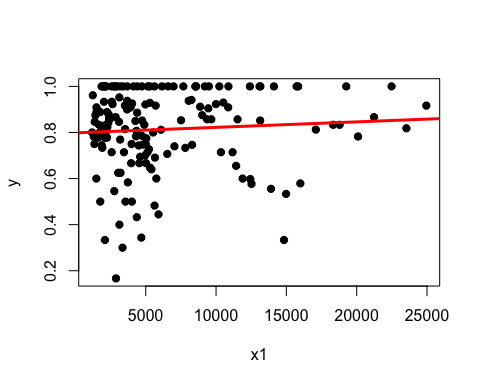
Interestingly, the plot below shows that the coastal ratio of the closest coastal state most likely does have a positive, significant effect on attack success, however the Philippines is a strong outlier.



When we add GDP per capita nothing is significant.

##   
## Call:  
## lm(formula = shipping$`success Ratio` ~ shipping$`attacks/Year` +   
## shipping$`coast/Area ratio (m/km2)` + shipping$`GDP per cap`,   
## data = shipping)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.54328 -0.06774 0.01918 0.12823 0.31843   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 8.613e-01 2.558e-02 33.674 <2e-16  
## shipping$`attacks/Year` -9.047e-04 4.821e-04 -1.877 0.0625  
## shipping$`coast/Area ratio (m/km2)` -1.504e-03 3.417e-04 -4.400 2e-05  
## shipping$`GDP per cap` 1.304e-06 2.581e-06 0.505 0.6140  
##   
## (Intercept) \*\*\*  
## shipping$`attacks/Year` .   
## shipping$`coast/Area ratio (m/km2)` \*\*\*  
## shipping$`GDP per cap`   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1649 on 155 degrees of freedom  
## Multiple R-squared: 0.1301, Adjusted R-squared: 0.1132   
## F-statistic: 7.726 on 3 and 155 DF, p-value: 7.641e-05

A graph of the above regression



Number of attacks is significant again.

## Oneway (individual) effect Within Model  
##   
## Call:  
## plm(formula = shipping$`success Ratio` ~ shipping$`attacks/Year` +   
## shipping$`GDP per cap`, data = shipping, model = "within",   
## index = c("country", "year"))  
##   
## Unbalanced Panel: n=8, T=17-21, N=159  
##   
## Residuals :  
## Min. 1st Qu. Median 3rd Qu. Max.   
## -0.6370 -0.0627 0.0247 0.0965 0.3100   
##   
## Coefficients :  
## Estimate Std. Error t-value Pr(>|t|)   
## shipping$`attacks/Year` -2.2823e-03 7.9659e-04 -2.8651 0.004772 \*\*  
## shipping$`GDP per cap` 9.3488e-06 5.4794e-06 1.7062 0.090062 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Total Sum of Squares: 3.9439  
## Residual Sum of Squares: 3.6638  
## R-Squared : 0.071044   
## Adj. R-Squared : 0.066576   
## F-statistic: 5.69754 on 2 and 149 DF, p-value: 0.0041271

Here are diagnostic regressions

#####################  
# Regression Diagnostics  
######################  
  
#Assessing homoscedasticity (we have met the constant variance assumption if p < 1.95)  
ncvTest(ols2)

## Non-constant Variance Score Test   
## Variance formula: ~ fitted.values   
## Chisquare = 3.406175 Df = 1 p = 0.06495285

#Assessing multicollinearity  
vif(ols2)

## shipping$`attacks/Year` shipping$`coast/Area ratio (m/km2)`   
## 1.001364 1.001364

#Assessing outliers  
outlierTest(ols2)

##   
## No Studentized residuals with Bonferonni p < 0.05  
## Largest |rstudent|:  
## rstudent unadjusted p-value Bonferonni p  
## 35 -3.363319 0.00097039 0.15429

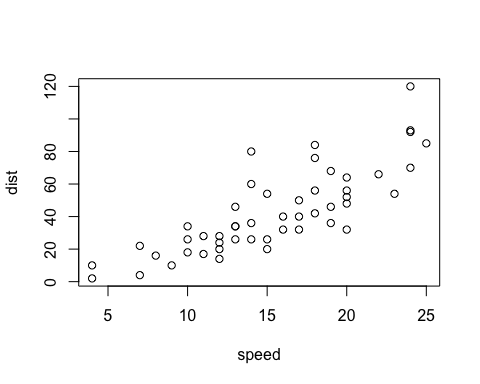
# Data gathering

We

summary(cars)

## speed dist   
## Min. : 4.0 Min. : 2.00   
## 1st Qu.:12.0 1st Qu.: 26.00   
## Median :15.0 Median : 36.00   
## Mean :15.4 Mean : 42.98   
## 3rd Qu.:19.0 3rd Qu.: 56.00   
## Max. :25.0 Max. :120.00

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.