

Advanced Data Analysis: Exam 1

Andrew Resnikoff

aresniko

Introduction

Background

As times change, ideas change with it. As new ideas are put forward, it is often difficult to change the minds of those around you. This analysis will look at many different political movements and how whether or not the movement was non violent ultimately affected the end result. Through this investigation, we aim to determine whether non violent movements are more likely to be successful ones and which factors are significant in determining whether violent and non violent movement will be successful.

Data

We are given a dataset with 323 observations of the following variables: whether the movement was successful, partially successful or a failure (response), the name, country and year of the movement, whether the movement was nonviolent, a measure of how democratic the government the movement takes place in was, whether the government was under sanctions or was recieving aid, whether foreign governments supported the movement, whether the government used violence to repress the movement, whether military and police forces sided with the movement and the duration of the movement in days.

We recode the outcome variable (whether the movement was successful) to be successful even when it has been classified as a partial success. We also remove all observations that contain NA, which leaves us with 267 complete observations.

Models

We begin this investigation by fitting a general additive model to our dataset. The variables campaign and country are removed by default. The variables year, democracy and duration are all smoothed using splines.

General Additive Model

```
gam.model<-gam(outcome~s(year)+nonviol+s(democracy)
               +sanctions+aid+support+viol.repress
               +defect+s(duration),data=navc.clean,family="binomial")
```

```
##
## Family: binomial
## Link function: logit
##
## Formula:
## outcome ~ s(year) + nonviol + s(democracy) + sanctions + aid +
##          support + viol.repress + defect + s(duration)
##
## Parametric coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.20046    0.52472  -0.382  0.70243
## nonviol       2.16926    0.40738   5.325 1.01e-07 ***
## sanctions     0.04006    0.44500   0.090  0.92827
## aid          -0.01863    0.32954  -0.057  0.95491
## support       0.75854    0.39706   1.910  0.05609 .
## viol.repress -0.89221    0.51643  -1.728  0.08405 .
## defect        1.01459    0.36886   2.751  0.00595 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##              edf Ref.df Chi.sq p-value
## s(year)        1.000  1.000  8.570 0.00342 **
## s(democracy)   5.574  6.653 14.439 0.03688 *
## s(duration)    1.000  1.000  2.139 0.14358
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.275   Deviance explained = 25.8%
## UBRE = 0.13289   Scale est. = 1          n = 267
```

Table 1: Model Summary

Variable Name	Coefficient	Standard Error	P-Value
(intercept)	-.200	.525	.702
nonviol	2.169	.407	0
sanctions	.040	.445	.928
aid	-.019	.330	.995
support	.759	.397	.056

Variable Name	Coefficient	Standard Error	P-Value
viol.repress	-.892	.516	.084
defect	1.015	.369	.006

Binary Predictor Variables

Since the coefficients of the non smoothed terms are essentially the effects of a logistic regression on these variables, we can summarize the effects of each of these terms holding all others constant.

nonviol A coefficient of 2.169 implies that a movement that uses nonviolence is $e^{2.169} = 8.75$ times more likely to succeed than a movement that is violent.

sanctions A coefficient of .4 implies that a movement under a government under international sanctions is $e^{.4} = 1.492$ times more likely to succeed than one that is not.

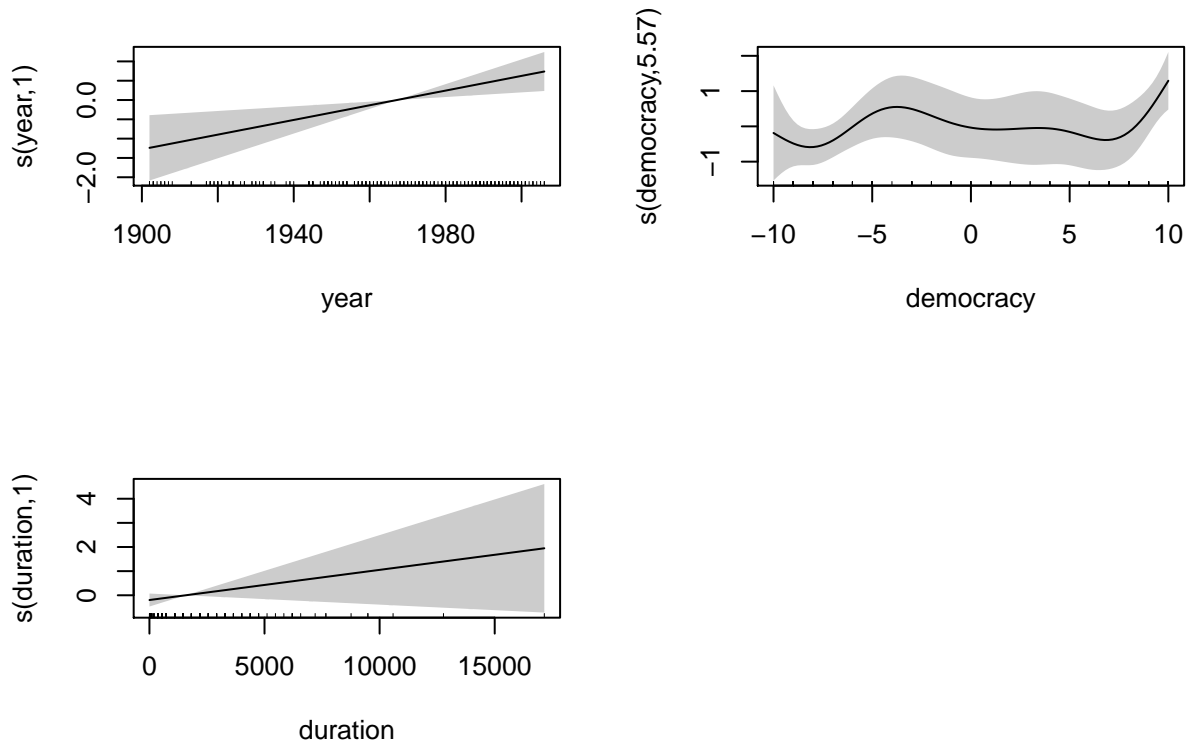
aid A coefficient of -.019 implies that a movement under a government who recieved aid from other countries to deal with the movement is $e^{-.019} = .981$ times as likely to succeed than one that is not.

support A coefficient of .759 implies that a movement that recieved aid from foreign governments is $e^{.759} = 2.136$ times as likely to succeed than one that is not.

viol.repress A coefficient of -.892 implies that a movement that is met with violent repression from the government is $e^{-.892} = .410$ times as likely to succeed than one that is not.

defect A coefficient of 1.015 implies that a movement which is supported by a substantial portion of the police and military is $e^{1.015} = 2.759$ times as likely to succeed as one that is not.

Continuous Predictor Variables



We can assess the relationship between the continuous predictor variables and the response by looking at the plots of the smoothing splines used in our model.

year The spline function estimated for the variable year has 1 effective degree of freedom and looks to be linear. This implies that as the year increases, movements are in general more likely to succeed. It is important to note that the spline takes on negative values and positive values, meaning that movements taking place before a certain year (around 1960 or 1970) were less likely to succeed and movements taking place after this year are more likely to succeed.

duration The spline function estimated for the variable duration has 1 effective degree of freedom and looks to be linear. This implies that the more days a movement goes on for, the more likely it is to survive. Since the value of the spline function is always positive, this means that a movement does not have to be going for a certain number of days before it becomes more likely to be successful.

democracy The spline function estimated for the variable democracy has 5.57 effective degrees of freedom and is not linear. The function oscillates, so different values for democracy have very different effects on the outcome.

Violence vs. Non Violence

To get a better understanding of how the outcome is affected by a movement being violent or non-violent, we can build models on the separated datasets and see how they differ.

```
violent<-navc.clean[which(navc.clean$nonviol == 0),]
nonviolent<-navc.clean[which(navc.clean$nonviol == 1),]

gam.violent<-gam(outcome~s(year)+s(democracy)
                 +sanctions+aid+support+viol.repress
                 +defect+s(duration),data=violent,family="binomial")

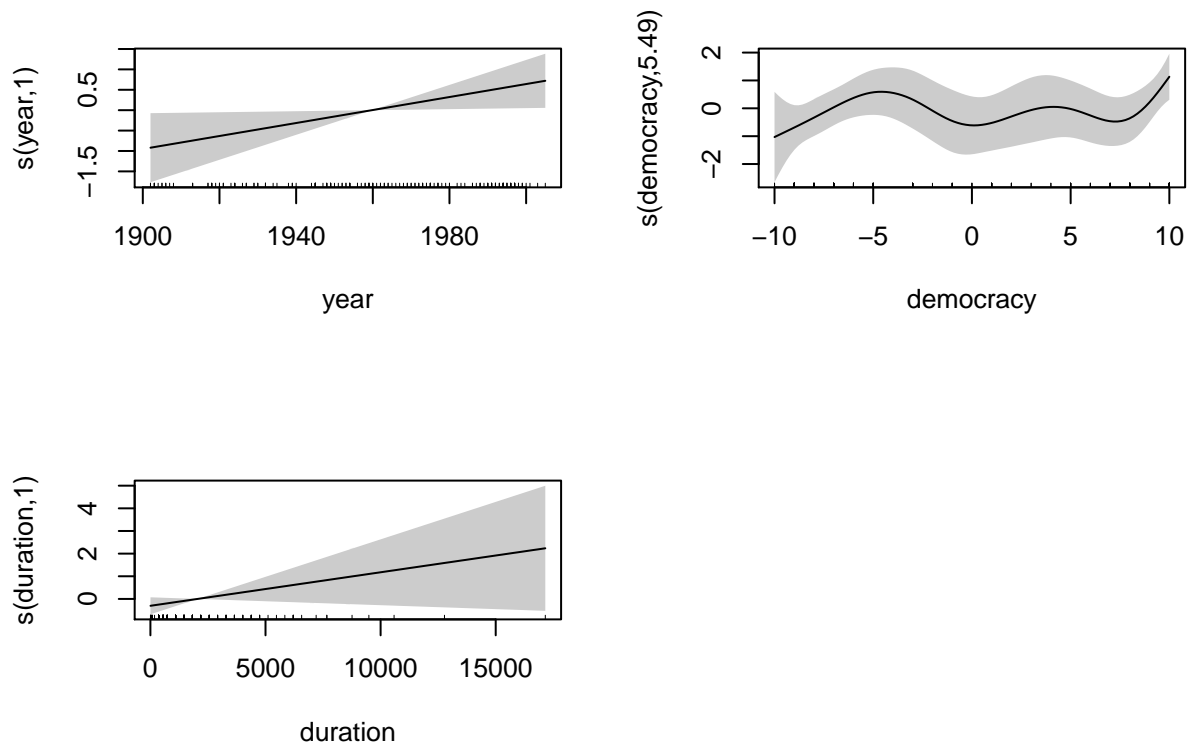
gam.nonviolent<-gam(outcome~s(year)+s(democracy)
                    +sanctions+aid+support+viol.repress
                    +defect+s(duration),data=nonviolent,family="binomial")
```

Violent Data Set Summary

```
##
## Family: binomial
## Link function: logit
##
## Formula:
## outcome ~ s(year) + s(democracy) + sanctions + aid + support +
##          viol.repress + defect + s(duration)
##
## Parametric coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.34362    0.62493  -0.550   0.582
## sanctions    0.64565    0.53281   1.212   0.226
## aid         -0.04259    0.40836  -0.104   0.917
## support      0.63850    0.42905   1.488   0.137
## viol.repress -0.73598    0.63305  -1.163   0.245
## defect       0.67281    0.43221   1.557   0.120
##
## Approximate significance of smooth terms:
##              edf Ref.df Chi.sq p-value
## s(year)        1.00  1.000  4.722  0.0298 *
```

```
## s(democracy) 5.49  6.551 12.486  0.0694 .
## s(duration)   1.00  1.000  2.622  0.1054
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.15   Deviance explained = 17.3%
## UBRE = 0.26275   Scale est. = 1           n = 172
```

Violent Data Set Plot

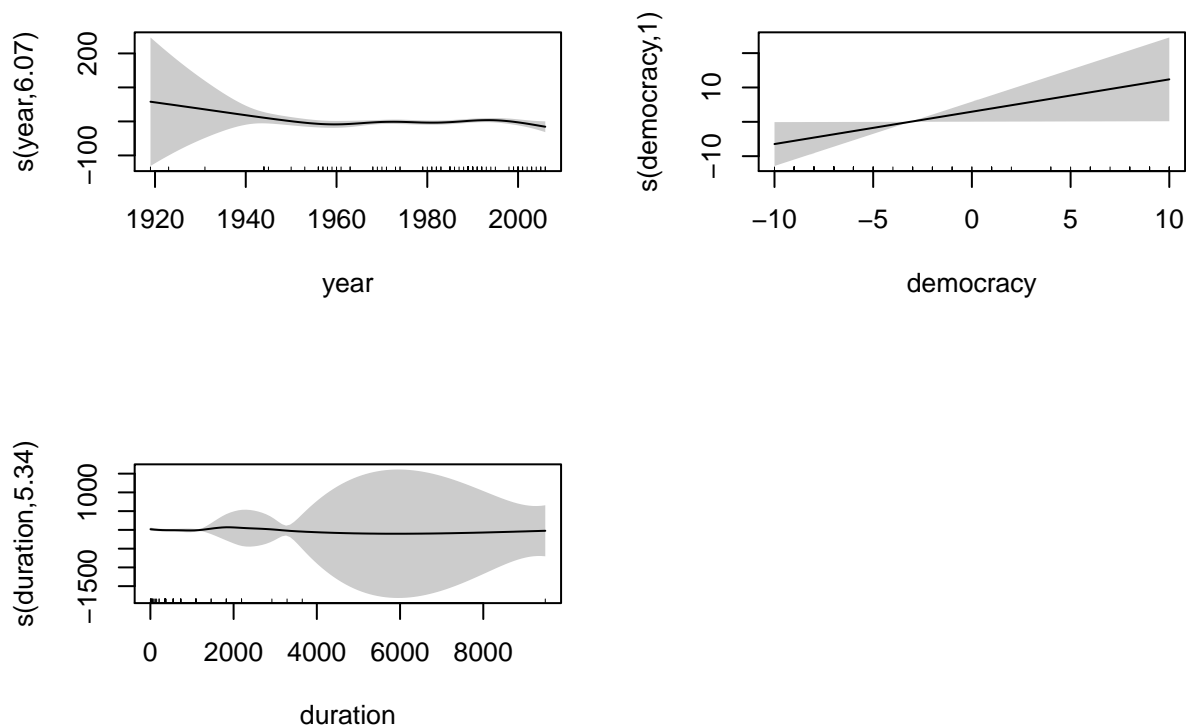


Non Violent Data Set Summary

```
##
## Family: binomial
## Link function: logit
##
## Formula:
## outcome ~ s(year) + s(democracy) + sanctions + aid + support +
##          viol.repress + defect + s(duration)
##
## Parametric coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   23.699    18.146   1.306  0.1915
## sanctions     -9.156     6.601  -1.387  0.1654
## aid           -1.868     1.616  -1.156  0.2477
## support        4.145     8.330   0.498  0.6188
## viol.repress  -9.782     7.037  -1.390  0.1645
## defect         8.655     4.160   2.080  0.0375 *
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##              edf Ref.df Chi.sq p-value
## s(year)      6.065  6.742 10.689  0.1372
## s(democracy) 1.000  1.000  4.111  0.0426 *
## s(duration)  5.336  5.607  1.932  0.9037
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.815   Deviance explained = 84.2%
## UBRE = -0.45965   Scale est. = 1          n = 95
```

Non Violent Data Set Plot



Looking at these two models, we can see many discrepancies in the way the two models behave. In violent movements, the government being under sanctions during the movement improves the chances of success, while in non violent movements it decreases the chances. Being a more recent year or the movement taking a longer duration will also improve chances of success in violent movements. In non violent movements, the year and duration have a mostly constant effect (an increase year seems to decrease chances of success slightly). The effect of democracy is also very different. In violent movements, we see an effect very similar to the initial models. In non violent movements, we see a linear relationship between democracy and success. A non violent movement under a stronger democracy is more likely to be successful.

There are also some similarities between the two models. The government receiving aid from foreign government has a negative effect on the chance of success in both violent and non violent movements. Receiving aid from foreign governments and having the support of the police and military will increase the chances of success (although the effect is larger in non violent movements, but this may just be because of

the scale of other variables.) A government violently repressing a movement will also decrease the chances of success for both types of movements (again the effect is larger in non violent movements.)

Model Validation

Cross Validation

We can use cross validation to determine the percentage of correct guesses that our models make. This seems to be the best way to see how well our models actually perform against test data. Looking at the percentage of correct guesses our models make will tell us whether our models actually predict accurately.

```
cv.gam(navc.clean,"outcome~s(year)+nonviol+s(democracy)
      +sanctions+aid+support+viol.repress
      +defect+s(duration)",nrow(navc.clean))
```

```
## [1] 0.1991155
```

```
cv.gam(violent,"outcome~s(year)+s(democracy)
      +sanctions+aid+support+viol.repress
      +defect+s(duration)",nrow(violent))
```

```
## [1] 0.2264103
```

```
cv.gam(nonviolent,"outcome~s(year)+s(democracy)
      +sanctions+aid+support+viol.repress
      +defect+s(duration)",nrow(nonviolent))
```

```
## [1] 0.1546
```

Our cross-validated percent of correct guesses for the initial model is 19.91%.

Our cross-validated percent of correct guesses for the violent model is 22.64%.

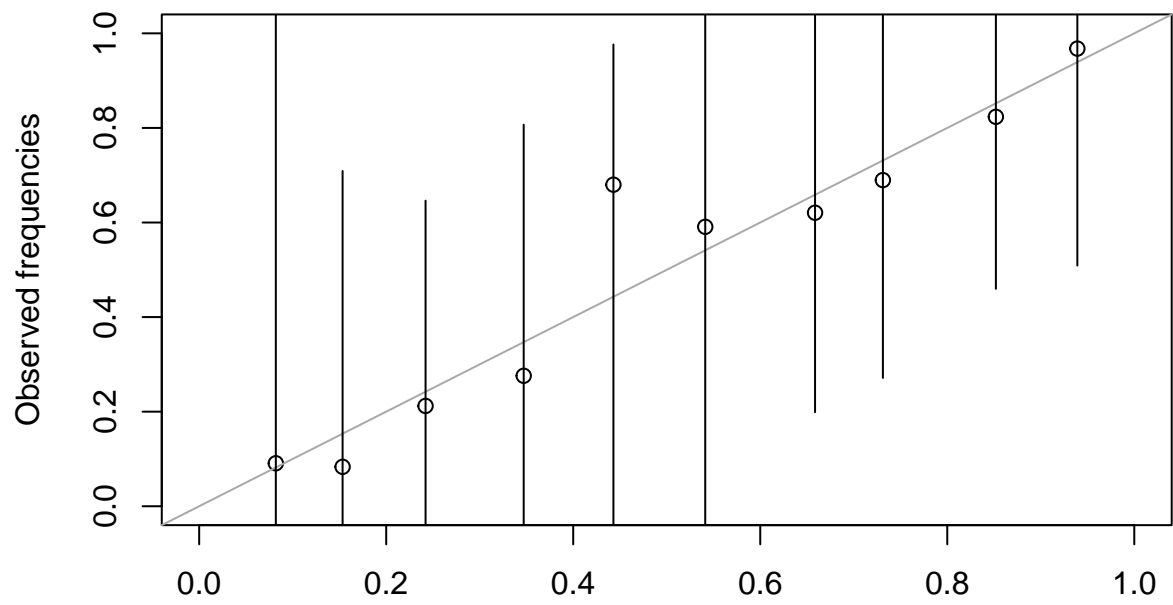
Our cross-validated percent of correct guesses for the non-violent model is 15.46%.

Notice that these numbers are very low in terms of the percentage they get correct. About 1 out of 5 predictions is correct when we build a model on all of the data. When we separate the data into violent and non violent movements, we do much better. On the whole dataset, we average about 53 total correct guesses. When violent and non violent movements are split, we average about 39 correct guesses on the violent data and about 14 correct guesses on the non violent data, a total of 54 correct guesses. While this isn't much better, it means that separating the data does help the models performance (especially for violent movements.)

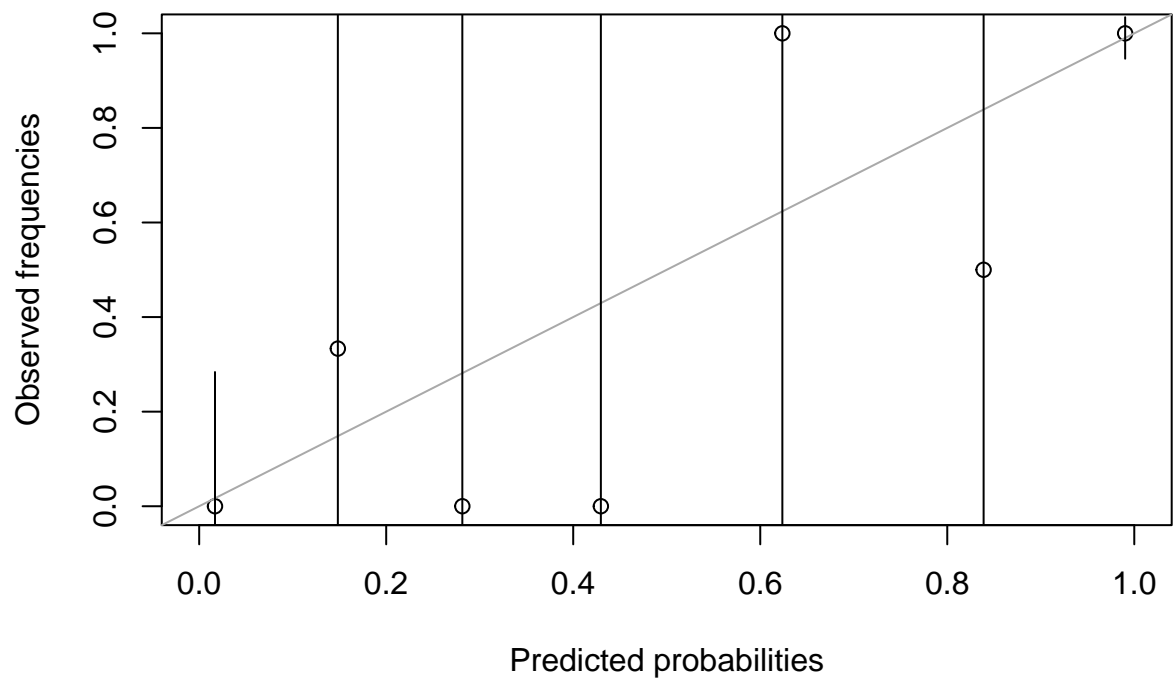
Observed Frequency Vs. Predicted Probability Plots

The observed frequency vs. predicted probability plots help us ensure that our models behave as expected. We expect that about $n\%$ of data observations to have their outcome correctly predicted when the predicted probability for those observations is $n\%$.

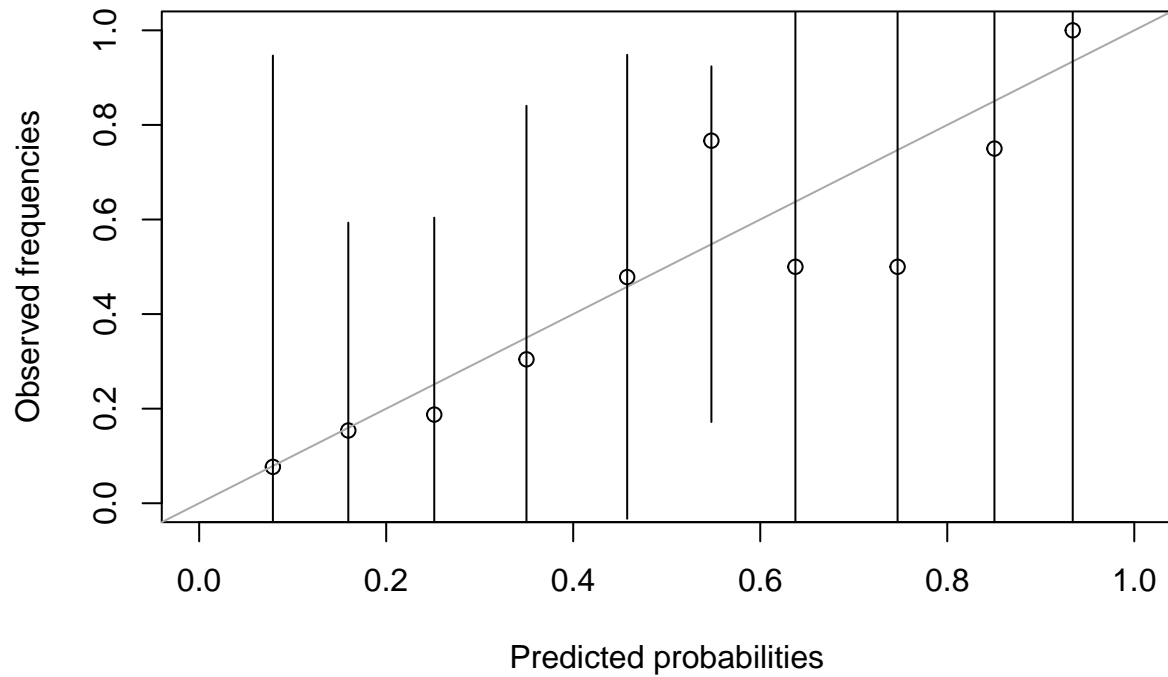
All Data



Non Violent Data



Violent Data



From the Observed frequencies vs. Predicted probabilities plots, we see that the models built on all of the data and the violent data behave as expected. The plot built from the non violent data, however, does not really line up as we would expect.

Final Model

It seems better in this case to move forward with the split model. While the Observed frequencies vs. Predicted probabilities plot for the non violent data does not seem to match up with our assumptions, the two separate models give us a much better idea as to what affects the outcome of a movement when it is violent and when it is nonviolent.

Results

From our initial model, it is evident that non violent movements are, in general, more successful than violent movements. This is made evident by the intercept coefficient of 23.699 in the nonviolent model. This value is the log odds of a non violent company being successful. Keep in mind, this implies that the probability of success for a nonviolent movement in general is about $e^{23.699} = 19604010508$. While we know that this is not a realistic probability, it speaks to the idea that being a non violent movement gives a much better chance of a successful outcome. Comparing this to the intercept coefficient of $e^{-0.34362} = .7091984$ for violent movements, despite the fact that the coefficients of these models are on different scales, we see this difference is very significant. When we add on the fact that in the initial model, we have a positive coefficient for the nonviol variable, we can conclude that non violent movements are more likely to be successful.

In both models (and the initial model), the coefficient of violent repression is negative. This implies that movements that are met with violent repression by their government are less likely to succeed.

The effect of aid and support on both violent and non violent movements is as you would expect. A government that recieves foreign aid to supress the movement in both types of movement decreases the chances of success of that movement. When the movement, and not the government, recieves that foreign aid instead, the movement becomes more likely. It is interesting to note that in both the violent and non violent models, the movement recieving foreign aid does more to increase the chances of a successful outcome than the government recieving foreign aid does to decrease the chances of a successful outcome (as suggested by the fact that the coefficient for support is more positive than the coefficient for aid is negative in both model summaries.)

From the plot of the estimated spline for the democracy variable in the non violent model, we see a clearly defined relationship between the level of a democracy of a government and the probability of success of a non violent momement. This relationship is linear, and appears to cross the line $y = 0$ at a democracy value of about -3. What this means, is that not only can we say that the more democractic a country is, the more likely that a non violent movement has a successful outcome but also that (holding all other factors constant) movements under a government with a democracy value of less than -3 will be less likely to succeed and movements under a government with a democracy value of more than -3 will be more likely to succeed (as when the estimate spline function takes a value of 0 it has no effect on the chances of success, and positive and negative values will increase and decrease chances of success respectively.)

In regards to how violent and non violent movements are affected by the level of democracy of the government, we see significant differences between the two types of movements. As explained above, in non violent movements, there is a positive linear relationship between the level of democracy and the chances of success. In violent movements, the relationship between democracy and chances of success oscillates. Since the estimated spline function is centered and close to zero, it seems like its possible that the chances of success for a violent movement don't really depend on how democratic the government is.