PS 733 methods paper

Hypothesis

The research studies the correlation between nationalist beliefs and patriarchal beliefs. The key variables are terms of nationalism and patriarchy, as well as authority in the interaction model. The hypothesis is that both patriarchy and authority should be significant and positively correlated with nationalism, and so is the interaction term of patriarchy and authority. The models also include other demographic terms for control, including age, sex, level of education, and upbringings in the rural or urban area.

Operationalizing variables

The data employed in the research is from East Asia Social Survey in 2008 about cultural and globalization in China, which includes measures of key concepts in the hypothesis—attitudes towards nationalism, patriarchy, as well as authority. All survey questions and choices are shown in part I of the appendix. The survey evaluates people's nationalist propensity with a series of question regarding people's agreement on nationalist statements (the specific questions are included in Appendix, graph 1). While the survey captures nationalistic attitudes on multi-dimensions—protectionism regarding policy, hawkish attitude towards national interests and cultural chauvinism—I choose hawkish attitude as the most intuitive indicator for nationalism in the general sense, which is in question L2 (graph 1): 'The country should following its national interests even if these would lead to conflicts with other nations.' Patriarchy and authority terms are measured in the same manner with different questions. The choices to the questions are in seven point Likert scale, including one for 'strongly agree,' two for 'agree,' three for 'somewhat agree,' four for 'neither agree nor disagree,' five for 'somewhat disagree,' six for 'disagree,' and seven for 'strongly disagree.' Instead of treating the survey results as ordered response, this paper recodes the dependent variable (nationalism) to be binary: '1' stands for being nationalistic, and '0' stands for not being nationalistic. The threshold for the divide is the self-reported agreement with nationalistic statement. In other words, all responses of one, two and three are recoded as '1' and four to seven as '0'. The other key variables are kept in the original scale as in the survey, ranging from '1' to '7,' and '1' is indicates strong propensity for agreement with patriarchy and authority.

DATA

```
load("D:/AY1S1/630/630 reserach/data/2008 eass/USED/3. DS0001/34607-0001-Data.rda")
CNdata <- da34607.0001[1:3010,c("V17","V18","V23","V24","V85","V86","V87","SEX","AGE","URBRURAL","DEGREE

library(prettyR)
f <- function(x,lbls){
  lbls <- sort(levels(x))
lbls <- (sub("^\\([0-9]+\\) +(.+$)", "\\1", lbls))
x<- as.numeric(sub("^\\(0*([0-9]+)\\).+$", "\\1",x))
x <- add.value.labels(x, lbls)
}

patr1 <- CNdata$V17
patr2 <- CNdata$V18
hier1 <- CNdata$V23
hier2 <- CNdata$V24
nati1 <- CNdata$V85
nati2 <- CNdata$V86
nati3 <- CNdata$V87
sex <- CNdata$SEX
age <-CNdata$AGE
urru <- CNdata$URBRURAL
edu <- CNdata$DEGREE
```

```
datanew <- data.frame(f(patr1),f(patr2),f(hier1),f(hier2),f(nati1),f(nati2),f(nati3),f(sex),age,f(urru)
```

```
## More value labels than values, only the first 4 will be used
```

```
datanew$nationalism <- datanew$f.nati2.
datanew$hierarchy <- datanew$f.hier1.+datanew$f.hier2.
datanew$sex <- datanew$f.sex.-1
datanew$patriarchy <- datanew$f.patr1.+ datanew$f.patr2.

modeldata <- datanew[,9:15]
modeldata$bination <- ifelse(modeldata$nationalism < 3, modeldata$bination <- 1,modeldata$bination <- 0)
names(modeldata)[2]<-"urru"
names(modeldata)[3]<-"edu"
#in this case, the threshold of the measurement for nationalism is set at 4, any answer below, includin
```

TESTING CORRELATION BETWEEN VARAIBLES

```
datatest <- modeldata[,1:7]
cortst <- cor(datatest,use="pairwise.complete.obs")
```

Model

As the result of the regression is binary, the distribution is binomial and therefore we employ logit model to derive the likelihood function. The first model includes key explanatory variable of patriarchy and also demographic control variables include age, sex, education level, urban/rural upbringings. These variables are chosen because they are all relevant to the dependent variable, based on literature. Nevertheless, they are also likely to be correlated with the independent variable, patriarchy. After forming a correlation matrix, it seems all demographic variables are correlated with patriarchy with no more than 0.3 of correlation. We therefore include all these variables in the model. As can been seen from the result, the explanatory variable is statistically significant on all levels and control variable sex is also significant on the 0.5 level. The direction of the parameter is also the same as the hypothesis: every one unit of decrease in the agreement with the patriarchal statement, there is 0.14 more probability for the respondents to be nationalist. Also men are more likely to be nationalist than women.

MODEL1

```
m1 <- glm(bination~patriarchy+urru+age+sex+edu,data=modeldata, family = binomial)

summary(m1)
```

```
##
## Call:
## glm(formula = bination ~ patriarchy + urru + age + sex + edu,
##     family = binomial, data = modeldata)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7578  -1.3769   0.7986   0.8819   1.4343
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.4839883  0.2693987   5.509 3.62e-08 ***
```

```
## patriarchy  -0.1396834  0.0175033  -7.980 1.46e-15 ***
## urru         0.0085192  0.0425059   0.200   0.8411
## age         -0.0008171  0.0030331  -0.269   0.7876
## sex         -0.1899806  0.0808631  -2.349   0.0188 *
## edu          0.0233789  0.0359283   0.651   0.5152
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3782.2  on 2981  degrees of freedom
## Residual deviance: 3708.6  on 2976  degrees of freedom
##   (28 observations deleted due to missingness)
## AIC: 3720.6
##
## Number of Fisher Scoring iterations: 4
```

```
#model with patriarchy as the key explanatory varaible, with demographic control varaiables inluding ag
BIC(m1)
```

```
## [1] 3756.572
```

In order to figure out the model fit, I calculated the predicted probability with the model and it seems the model can account for above 60% percent of the data. The parentage changes as we adjust the threshold for '1'outcome and '0' outcome. There is, however, not much change between setting the threshold at 0.5 and 0.6, but trade-off between type I error and type II error.

MODEL FIT

```
modeldata$predvalue <- predict(m1,modeldata, type="response")

a<-ifelse(modeldata[,9] < 0.5, 0, 1)
modeldata<- cbind(modeldata,a)
modeldata$fit <- NA

for (i in 1:3010){
   ifelse (modeldata[i,8]==modeldata[i,10],modeldata[i,11]<-1,modeldata[i,11]<-0)}
summary(modeldata$fit)
```

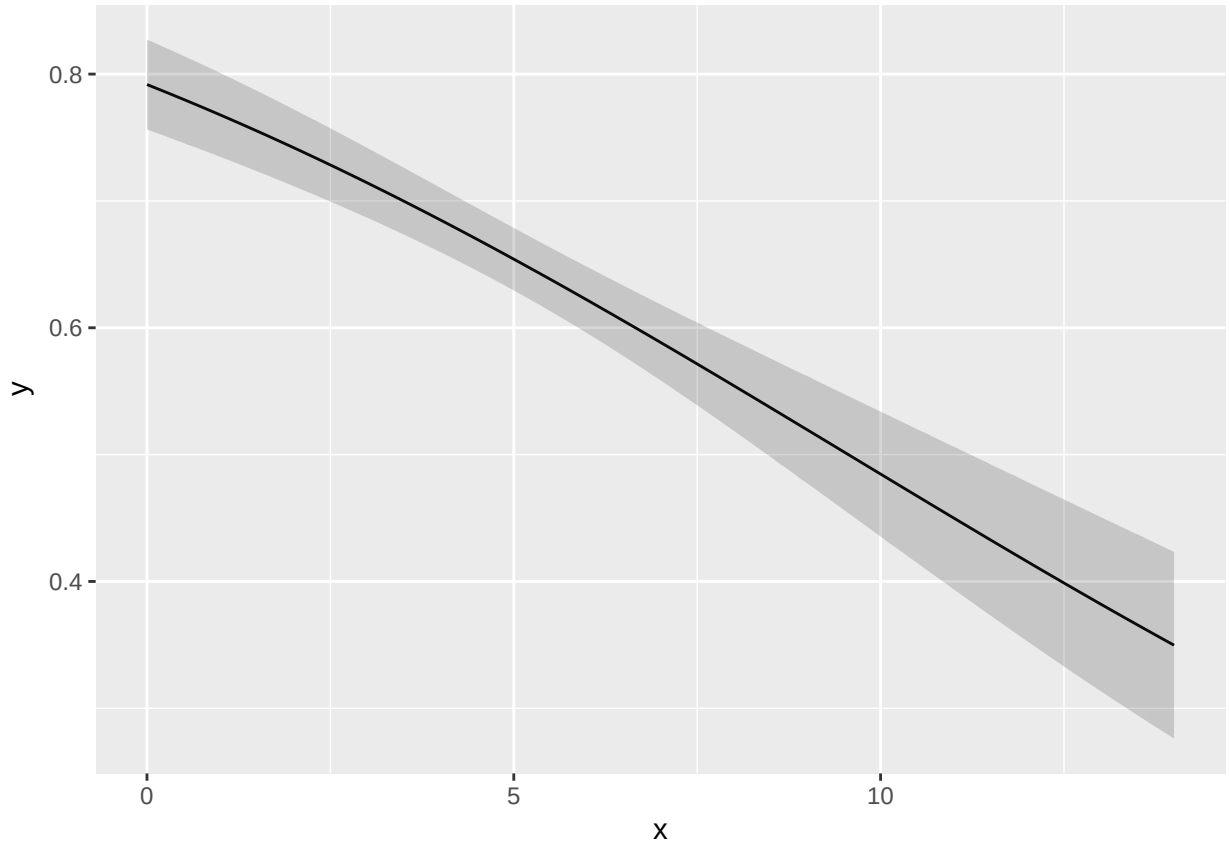
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   0.000   0.000   1.000   0.665   1.000   1.000      28
```

```
#mean=0.6761, the model predict about 67.6% of the outcome correctly. threshold at 0.5.
# 0.665, when set threshold at 0.6. No big difference, only trade off between type I and type II error.
```

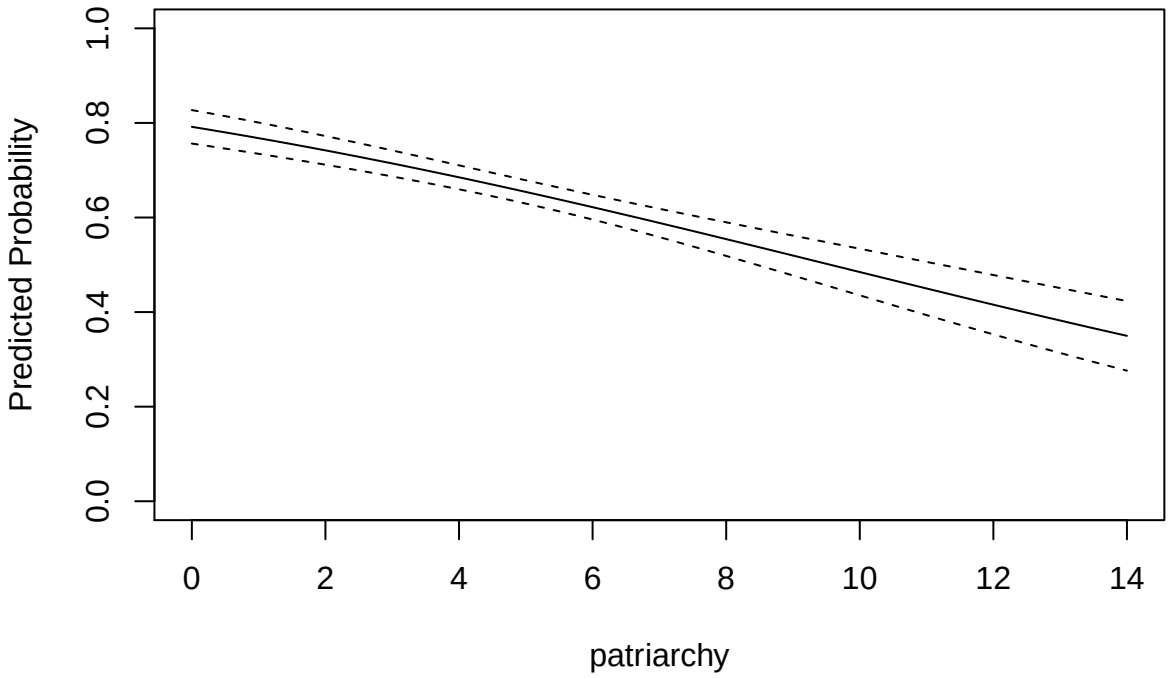
representation of uncertainty with interval differences using standard error derived from the model

```
aveage <- mean(modeldata$age)
aveedu <- mean(modeldata$edu,na.rm = T)
aveurru <- mean(modeldata$urru,na.rm = T)

datapred <- data.frame(edu <-rep(aveedu,140),age <-rep(aveage,140),sex<-rep(median(modeldata$sex),140),
```

```r
names(datapred)[1]<- "edu"
names(datapred)[2]<- "age"
names(datapred)[3]<- "sex"
names(datapred)[4]<- "urru"
names(datapred)[5]<- "patriarchy"
datapred <- na.omit(datapred)
a<-predict(m1,datapred,type='response',se.fit = TRUE)
a<-data.frame(x=datapred$patriarchy,a)

hi<-a$fit+2*a$se.fit
lo<-a$fit-2*a$se.fit

c<-data.frame(y=a$fit,hi=hi,lo=lo,x=a$x)
library(ggplot2)
ggplot(c,aes(x=x,y=y,ymin=lo,ymax=hi))+geom_line()+geom_ribbon(alpha=.2)
```




```r
plot(a$x,a$fit, type="l", ylim=c(0,1), xlab="patriarchy", ylab="Predicted Probability")
lines(a$x,hi,lty=2)
lines(a$x,lo,lty=2)
```



INTERACTION MODEL

To formulate the interaction model, I first tested the interaction between variables using the Akaike information criterion, in order to figure out the available terms for interaction. As is seemed, interaction between sex and patriarchy is both theoretically correct and recommended by Akaike information criterion. A test of correlation is performed and yield a result of correlation of 0.029. There interaction model is therefore formed as in table 2. There is not much change in the coefficient for patriarchy in terms of its significance or value. Sex increases in significance and the interaction term is also significant on a 0.5 level.

EVALUATE INTERACTION TERMS

```
a <- glm(bination~(patriarchy+urru+age+sex+edu)^2,data= modeldata)
step(a,k=2)
```

```
## Start:  AIC=3907.66
## bination ~ (patriarchy + urru + age + sex + edu)^2
##
##                    Df Deviance    AIC
## - urru:sex          1   640.02 3905.7
## - sex:edu           1   640.02 3905.7
## - patriarchy:age    1   640.05 3905.8
## - patriarchy:edu    1   640.06 3905.9
## - age:sex           1   640.08 3906.0
## - urru:edu          1   640.12 3906.1
## - urru:age          1   640.17 3906.4
## - patriarchy:urru   1   640.17 3906.4
## <none>                  640.01 3907.7
```

```
## - age:edu          1   640.56 3908.2
## - patriarchy:sex   1   641.03 3910.4
##
## Step:  AIC=3905.68
## bination ~ patriarchy + urru + age + sex + edu + patriarchy:urru +
##     patriarchy:age + patriarchy:sex + patriarchy:edu + urru:age +
##     urru:edu + age:sex + age:edu + sex:edu
##
##                   Df Deviance    AIC
## - sex:edu          1   640.02 3903.7
## - patriarchy:age   1   640.05 3903.9
## - patriarchy:edu   1   640.06 3903.9
## - age:sex          1   640.09 3904.0
## - urru:edu         1   640.13 3904.2
## - urru:age         1   640.17 3904.4
## - patriarchy:urru  1   640.17 3904.4
## <none>                 640.02 3905.7
## - age:edu          1   640.57 3906.3
## - patriarchy:sex   1   641.05 3908.5
##
## Step:  AIC=3903.7
## bination ~ patriarchy + urru + age + sex + edu + patriarchy:urru +
##     patriarchy:age + patriarchy:sex + patriarchy:edu + urru:age +
##     urru:edu + age:sex + age:edu
##
##                   Df Deviance    AIC
## - patriarchy:age   1   640.06 3901.9
## - patriarchy:edu   1   640.06 3901.9
## - age:sex          1   640.12 3902.2
## - urru:edu         1   640.13 3902.2
## - urru:age         1   640.17 3902.4
## - patriarchy:urru  1   640.17 3902.4
## <none>                 640.02 3903.7
## - age:edu          1   640.58 3904.3
## - patriarchy:sex   1   641.08 3906.6
##
## Step:  AIC=3901.87
## bination ~ patriarchy + urru + age + sex + edu + patriarchy:urru +
##     patriarchy:sex + patriarchy:edu + urru:age + urru:edu + age:sex +
##     age:edu
##
##                   Df Deviance    AIC
## - patriarchy:edu   1   640.08 3900.0
## - urru:edu         1   640.16 3900.4
## - age:sex          1   640.16 3900.4
## - patriarchy:urru  1   640.20 3900.6
## - urru:age         1   640.22 3900.6
## <none>                 640.06 3901.9
## - age:edu          1   640.67 3902.7
## - patriarchy:sex   1   641.08 3904.6
##
## Step:  AIC=3899.97
## bination ~ patriarchy + urru + age + sex + edu + patriarchy:urru +
##     patriarchy:sex + urru:age + urru:edu + age:sex + age:edu
```

```
##
##                   Df Deviance    AIC
## - urru:edu         1   640.18 3898.4
## - age:sex          1   640.18 3898.5
## - patriarchy:urru  1   640.20 3898.6
## - urru:age         1   640.25 3898.8
## <none>                 640.08 3900.0
## - age:edu          1   640.67 3900.7
## - patriarchy:sex   1   641.08 3902.6
##
## Step:  AIC=3898.44
## bination ~ patriarchy + urru + age + sex + edu + patriarchy:urru +
##     patriarchy:sex + urru:age + age:sex + age:edu
##
##                   Df Deviance    AIC
## - age:sex          1   640.28 3896.9
## - patriarchy:urru  1   640.37 3897.3
## - urru:age         1   640.51 3898.0
## <none>                 640.18 3898.4
## - age:edu          1   640.75 3899.1
## - patriarchy:sex   1   641.20 3901.2
##
## Step:  AIC=3896.91
## bination ~ patriarchy + urru + age + sex + edu + patriarchy:urru +
##     patriarchy:sex + urru:age + age:edu
##
##                   Df Deviance    AIC
## - patriarchy:urru  1   640.47 3895.8
## - urru:age         1   640.65 3896.6
## <none>                 640.28 3896.9
## - age:edu          1   640.79 3897.3
## - patriarchy:sex   1   641.24 3899.4
##
## Step:  AIC=3895.81
## bination ~ patriarchy + urru + age + sex + edu + patriarchy:sex +
##     urru:age + age:edu
##
##                  Df Deviance    AIC
## <none>                640.47 3895.8
## - urru:age        1   640.92 3895.9
## - age:edu         1   640.97 3896.1
## - patriarchy:sex  1   641.41 3898.2


##
## Call:  glm(formula = bination ~ patriarchy + urru + age + sex + edu +
##     patriarchy:sex + urru:age + age:edu, data = modeldata)
##
## Coefficients:
##    (Intercept)      patriarchy            urru             age
##      0.8398308      -0.0394180       0.0390118       0.0009074
##            sex             edu  patriarchy:sex        urru:age
##     -0.1203209      -0.0245121       0.0156754      -0.0008700
##        age:edu
##      0.0006888
```

```
##
## Degrees of Freedom: 2981 Total (i.e. Null);  2973 Residual
##   (28 observations deleted due to missingness)
## Null Deviance:       659.3
## Residual Deviance: 640.5     AIC: 3896
```

choose interactive model considering theory

```
#correlation test for interaction model
cor.test(modeldata$sex,modeldata$patriarchy,data= modeldata)
```

```
##
##  Pearson's product-moment correlation
##
## data:  modeldata$sex and modeldata$patriarchy
## t = 1.6271, df = 3007, p-value = 0.1038
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.006079976  0.065323103
## sample estimates:
##       cor
## 0.0296594
```

```
m2 <- glm(bination~sex*patriarchy+urru+age+sex+edu,family = binomial, data = modeldata)
summary(m2)
```

```
##
## Call:
## glm(formula = bination ~ sex * patriarchy + urru + age + sex +
##     edu, family = binomial, data = modeldata)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8145  -1.3728   0.7921   0.8959   1.5292
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)     1.6966562  0.2870722   5.910 3.42e-09 ***
## sex            -0.5843184  0.1957689  -2.985  0.00284 **
## patriarchy     -0.1808484  0.0256884  -7.040 1.92e-12 ***
## urru            0.0073849  0.0425350   0.174  0.86216
## age            -0.0006676  0.0030373  -0.220  0.82603
## edu             0.0225885  0.0359562   0.628  0.52986
## sex:patriarchy  0.0756535  0.0341637   2.214  0.02680 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3782.2  on 2981  degrees of freedom
## Residual deviance: 3703.6  on 2975  degrees of freedom
##   (28 observations deleted due to missingness)
## AIC: 3717.6
```

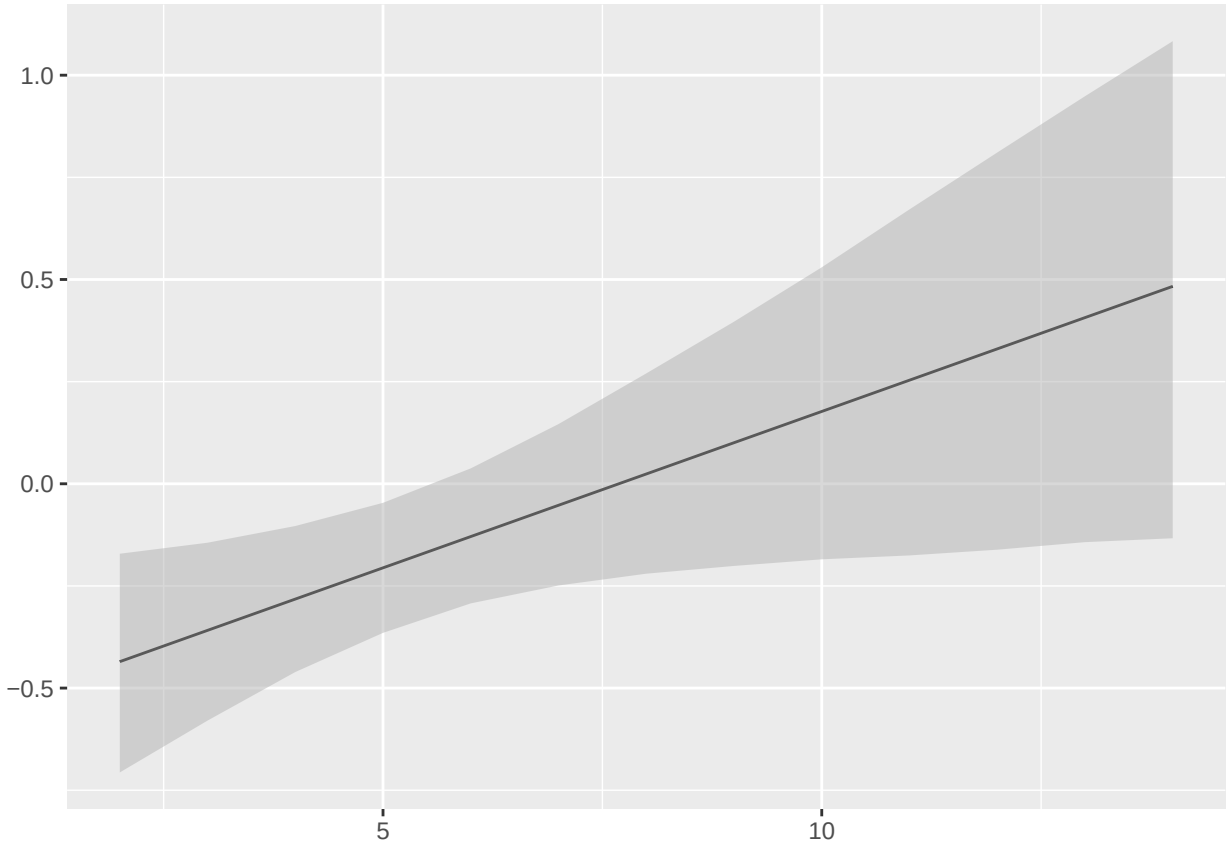
```
##
## Number of Fisher Scoring iterations: 4
```

UNCERTAINTY

```
library(interplot)
```

```
## Warning: package 'interplot' was built under R version 3.3.3

## Loading required package: abind

## Loading required package: arm

## Warning: package 'arm' was built under R version 3.3.3

## Loading required package: MASS

## Loading required package: Matrix

## Loading required package: lme4

## Warning: package 'lme4' was built under R version 3.3.3

##
## arm (Version 1.9-3, built: 2016-11-21)

## Working directory is C:/Users/Administrator/Desktop/733MLE/pp1
```

```
interplot(m2, var1 = "sex", var2 = "patriarchy",ci = 0.95)
```



Limitations

Most of the casual inference flaws come from the recoded data. It is nevertheless arguable that whether the variables can be treated as interval instead of ordinal. Most of the variables used in the model, except for age, can be treated as factor variable. Similar issues arise with treating categorized variables like educational level and upbringings, and I further compared a different model that treats the variable as factor instead of numeric. The binary nationalism term can arguably over-represent the degree of agreement with nationalism in the general population by equaling 'agree' with 'strongly agree,' also because the result already has acquiescence bias since people has the tendency to agree than disagree. Nevertheless, as observed in the data, the median is 'agree' instead of often chosen middle terms like 'somewhat agree' or 'neither agree or disagree,' which suggest that the general population's attitude is not neutral and there is the general propensity towards nationalism. This is captured by the binary categorization, as demonstrated in the histogram between original data and treated data.

model modified by factor

```
mf <- glm(bination~factor(patriarchy)+factor(urru)+age+sex+factor(edu),data = modeldata, family = binom
summary(mf)
```

```
##
## Call:
## glm(formula = bination ~ factor(patriarchy) + factor(urru) +
##     age + sex + factor(edu), family = binomial, data = modeldata)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9583  -1.3136   0.7539   0.9113   1.6196
```

```
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)             1.2958069  0.3090860   4.192 2.76e-05 ***
## factor(patriarchy)3    -0.2337771  0.2013509  -1.161 0.245624
## factor(patriarchy)4    -0.6711605  0.1763990  -3.805 0.000142 ***
## factor(patriarchy)5    -1.0953948  0.1929207  -5.678 1.36e-08 ***
## factor(patriarchy)6    -1.0177958  0.2044687  -4.978 6.43e-07 ***
## factor(patriarchy)7    -1.2732787  0.2124646  -5.993 2.06e-09 ***
## factor(patriarchy)8    -1.3254273  0.2132337  -6.216 5.10e-10 ***
## factor(patriarchy)9    -0.7178856  0.3134702  -2.290 0.022014 *
## factor(patriarchy)10   -2.0579571  0.3274466  -6.285 3.28e-10 ***
## factor(patriarchy)11   -1.4380854  0.4371831  -3.289 0.001004 **
## factor(patriarchy)12   -0.6957693  0.3158089  -2.203 0.027585 *
## factor(patriarchy)13   -1.5491996  1.0164396  -1.524 0.127473
## factor(patriarchy)14   -1.3101537  0.9409318  -1.392 0.163801
## factor(urru)2           0.6339624  0.3230568   1.962 0.049718 *
## factor(urru)3          -0.0828994  0.1155378  -0.718 0.473060
## factor(urru)4           0.0218964  0.1329128   0.165 0.869146
## age                     0.0001296  0.0030940   0.042 0.966587
## sex                    -0.1592772  0.0822938  -1.935 0.052933 .
## factor(edu)1            0.2777403  0.1566515   1.773 0.076232 .
## factor(edu)2            0.3825560  0.1620749   2.360 0.018257 *
## factor(edu)3            0.4372877  0.1745242   2.506 0.012224 *
## factor(edu)4            0.1662737  0.2065767   0.805 0.420877
## factor(edu)5            0.4109497  0.2332618   1.762 0.078111 .
## factor(edu)6           -0.3256886  0.7098603  -0.459 0.646373
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3782.2  on 2981  degrees of freedom
## Residual deviance: 3641.8  on 2958  degrees of freedom
##   (28 observations deleted due to missingness)
## AIC: 3689.8
##
## Number of Fisher Scoring iterations: 4
```

The interaction model captures the interactive effect between the two terms that is both quantitative and qualitative: the interaction terms are not only theoretically based, but also proved to be better fit according to the Akaike information criterion. The problem comes back to the recoding: whether it is reasonable to measure the interaction by multiply two arguably categorical variables. This may further render some heteroscedasticity issue, comparing the confidence interval plot in the original model and interaction model. In conclusion, the hypothesis about both significance and direction of the coefficients are is confirmed.