Microeconometrics, Empirical project, Group 8

Atanasov Georgi* Fitter Jonathan† Geyer Niklas‡ Hochholzer Matthias§
Woharcik Verena¶

7th February 2021

Importing data

from Wooldridge, his source: J. Grogger (1991), "Certainty vs. Severity of Punishment," Economic Inquiry 29, 297-309.

```
df<-read.dta("http://fmwww.bc.edu/ec-p/data/wooldridge/crime1.dta")
attach(df)
head(df)
```

```
##   narr86 nfarr86 nparr86 pcnv avgsen tottime ptime86 qemp86 inc86 durat black
## 1      0       0       0 0.38   17.6    35.2      12      0   0.0     0     0
## 2      2       2       0 0.44    0.0     0.0       0      1   0.8     0     0
## 3      1       1       0 0.33   22.8    22.8       0      0   0.0    11     1
## 4      2       2       1 0.25    0.0     0.0       5      2   8.8     0     0
## 5      1       1       0 0.00    0.0     0.0       0      2   8.1     1     0
## 6      0       0       0 1.00    0.0     0.0       0      4  97.6     0     0
##   hispan born60 pcnvsq pt86sq    inc86sq
## 1      0      1 0.1444    144    0.00000
## 2      1      0 0.1936      0    0.64000
## 3      0      1 0.1089      0    0.00000
## 4      1      1 0.0625     25   77.44000
## 5      0      0 0.0000      0   65.61001
## 6      0      1 1.0000      0 9525.75977
```

```
str(df)
```

```
## 'data.frame':    2725 obs. of  16 variables:
##  $ narr86 : num  0 2 1 2 1 0 2 5 0 0 ...
##  $ nfarr86: num  0 2 1 2 1 0 2 3 0 0 ...
##  $ nparr86: num  0 0 0 1 0 0 1 5 0 0 ...
##  $ pcnv   : num  0.38 0.44 0.33 0.25 0 ...
##  $ avgsen : num  17.6 0 22.8 0 0 ...
##  $ tottime: num  35.2 0 22.8 0 0 ...
##  $ ptime86: num  12 0 0 5 0 0 0 0 9 0 ...
##  $ qemp86 : num  0 1 0 2 2 4 0 0 0 3 ...
##  $ inc86  : num  0 0.8 0 8.8 8.1 ...
##  $ durat  : num  0 0 11 0 1 ...
```

*student ID
†student ID 11709902
‡student ID
§student ID 11724853
¶student ID

```
##  $ black  : num  0 0 1 0 0 0 1 0 1 0 ...
##  $ hispan : num  0 1 0 1 0 0 0 0 0 1 ...
##  $ born60 : num  1 0 1 1 0 1 1 1 1 1 ...
##  $ pcnvsq : num  0.1444 0.1936 0.1089 0.0625 0 ...
##  $ pt86sq : num  144 0 0 25 0 0 0 0 81 0 ...
##  $ inc86sq: num  0 0.64 0 77.44 65.61 ...
##  - attr(*, "datalabel")= chr ""
##  - attr(*, "time.stamp")= chr "10 Jan 2000 16:54"
##  - attr(*, "formats")= chr  "%9.0g" "%9.0g" "%9.0g" "%9.0g" ...
##  - attr(*, "types")= int  102 102 102 102 102 102 102 102 102 102 ...
##  - attr(*, "val.labels")= chr  "" "" "" "" ...
##  - attr(*, "var.labels")= chr  "" "" "" "" ...
##  - attr(*, "version")= int 6
```

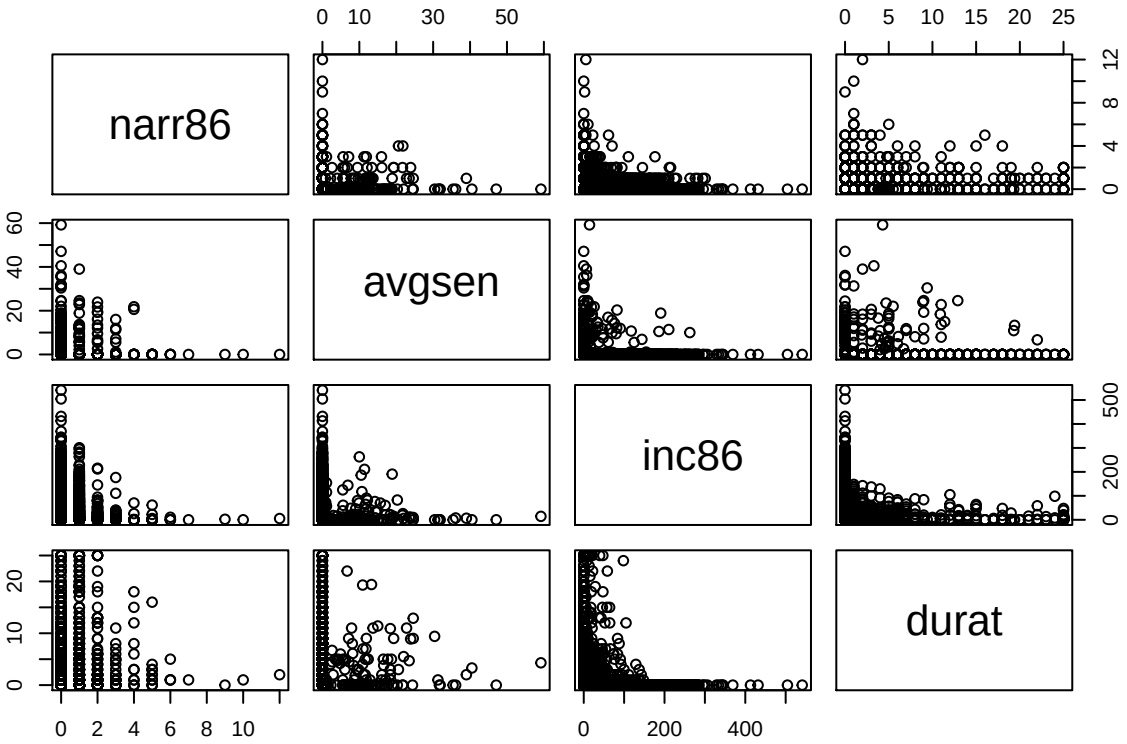
```
summary(df)
```

```
##      narr86           nfarr86          nparr86            pcnv
##  Min.   : 0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.: 0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median : 0.0000   Median :0.0000   Median :0.0000   Median :0.2500
##  Mean   : 0.4044   Mean   :0.2334   Mean   :0.1255   Mean   :0.3578
##  3rd Qu.: 1.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.6700
##  Max.   :12.0000   Max.   :6.0000   Max.   :8.0000   Max.   :1.0000
##      avgsen           tottime           ptime86            qemp86
##  Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000   Min.   :0.000
##  1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.:1.000
##  Median : 0.0000   Median : 0.0000   Median : 0.0000   Median :3.000
##  Mean   : 0.6323   Mean   : 0.8387   Mean   : 0.3872   Mean   :2.309
##  3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.:4.000
##  Max.   :59.2000   Max.   :63.4000   Max.   :12.0000   Max.   :4.000
##      inc86            durat            black            hispan
##  Min.   :  0.00   Min.   : 0.000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:  0.40   1st Qu.: 0.000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median : 29.00   Median : 0.000   Median :0.0000   Median :0.0000
##  Mean   : 54.97   Mean   : 2.251   Mean   :0.1611   Mean   :0.2176
##  3rd Qu.: 90.10   3rd Qu.: 2.000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :541.00   Max.   :25.000   Max.   :1.0000   Max.   :1.0000
##      born60           pcnvsq           pt86sq           inc86sq
##  Min.   :0.0000   Min.   :0.0000   Min.   :  0.000   Min.   :     0.00
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:  0.000   1st Qu.:     0.16
##  Median :0.0000   Median :0.0625   Median :  0.000   Median :   841.00
##  Mean   :0.3626   Mean   :0.2841   Mean   :  3.951   Mean   :  7458.93
##  3rd Qu.:1.0000   3rd Qu.:0.4489   3rd Qu.:  0.000   3rd Qu.:  8118.01
##  Max.   :1.0000   Max.   :1.0000   Max.   :144.000   Max.   :292681.00
```

A data.frame with 2725 observations on 16 variables: - narr86: times arrested, 1986 - nfarr86: felony arrests, 1986 - nparr86: property crme arr., 1986 - pcnv: proportion of prior convictions - avgsen: avg sentence length, mos. - tottime: time in prison since 18 (mos.) - ptime86: mos. in prison during 1986 - qemp86: quarters employed, 1986 - inc86: legal income, 1986, $100s - durat: recent unemp duration - black: =1 if black - hispan: =1 if Hispanic - born60: =1 if born in 1960 - pcnvsq: pcnv^2 - pt86sq: ptime86^2 - inc86sq: inc86^2

Correlation Plots

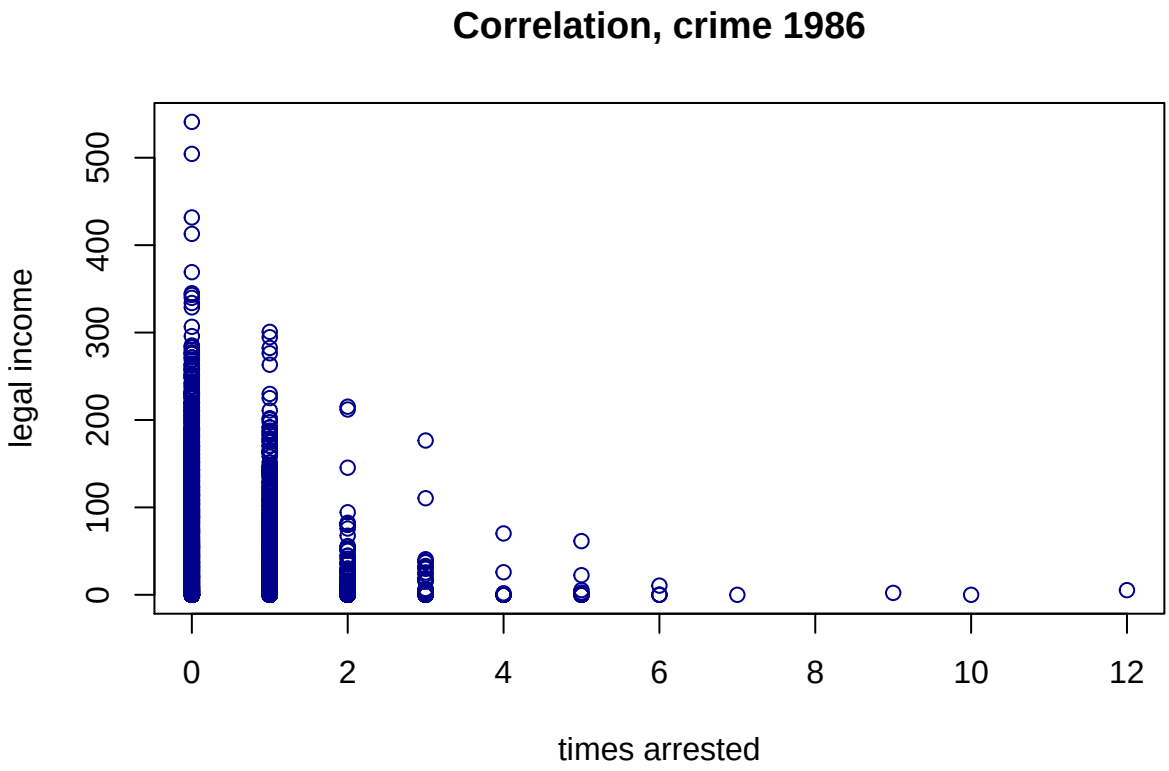
```
plot(df[,c("narr86", "avgsen", "inc86", "durat")])
```

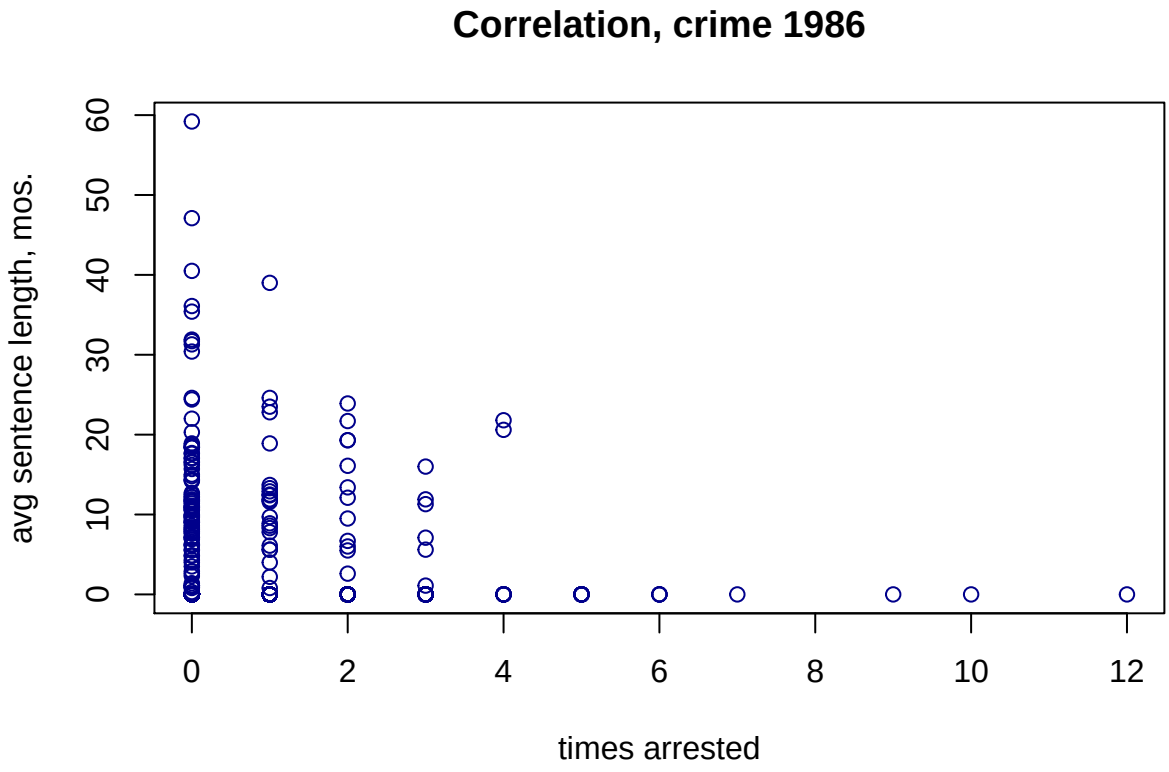


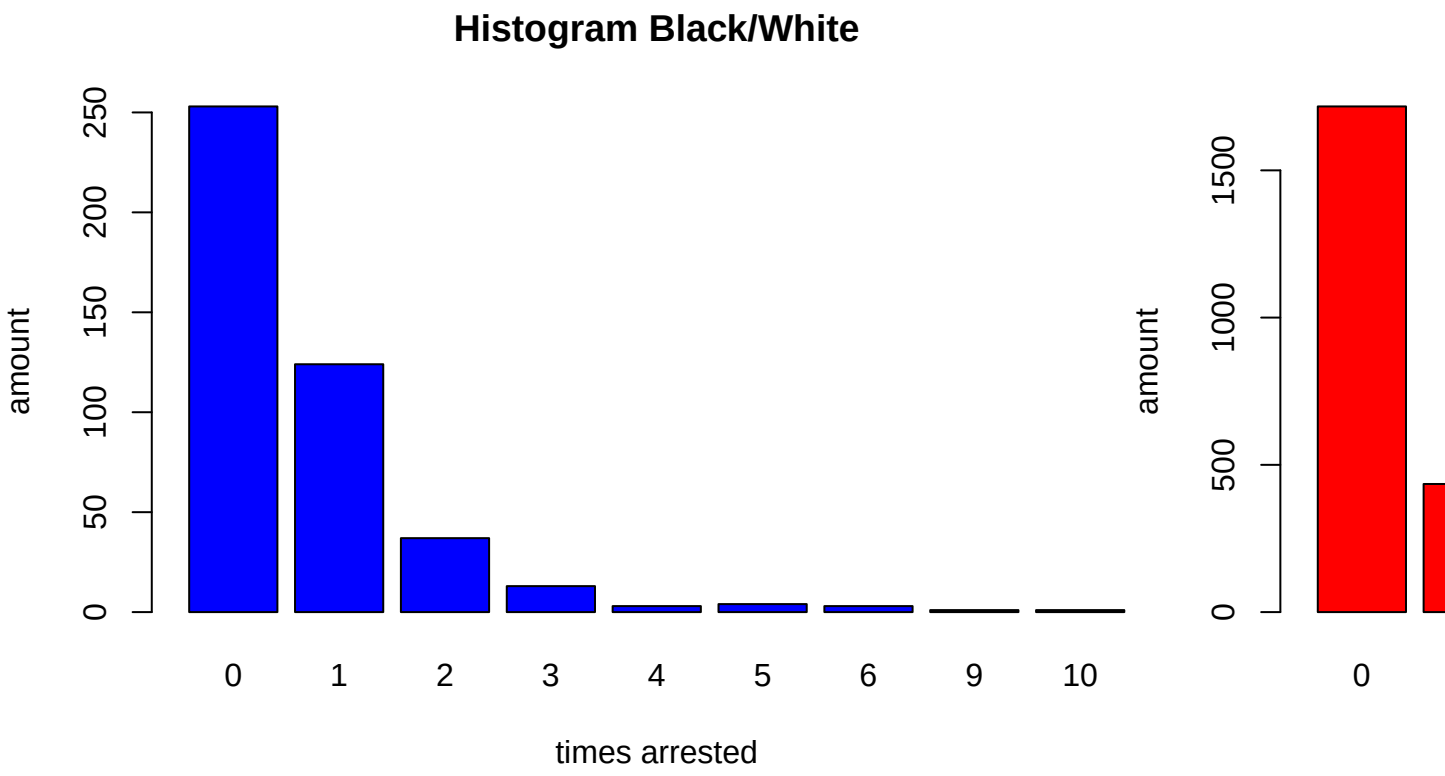

```
cor(df[,c("narr86", "avgsen", "inc86", "durat")])
```

```
##             narr86      avgsen       inc86       durat
## narr86  1.00000000  0.02929780 -0.18997653  0.08232769
## avgsen  0.02929780  1.00000000 -0.09580596  0.02843162
## inc86  -0.18997653 -0.09580596  1.00000000 -0.34292954
## durat   0.08232769  0.02843162 -0.34292954  1.00000000
```

Specific Plots:







############Georgi#################### ##############################

Working on "avgsen"

The idea is to build a model estimating expected severity of conviction when arrested in 1986 using level of income, employment, total time spend in prison and color (black ad non-black) of the arrested. Hypothesis, which are to be tested, are that the coefficient of every single variable is equal to 0.

Linear Model

The average severity is regressed on the above mentioned variables.

```
lm_sev2<-lm(avgsen~ tottime+ black+ qemp86+ inc86, data = df)
```

An output of an OLS-Estimation is given:

```
summary((lm_sev2))
```

```
##
## Call:
## lm(formula = avgsen ~ tottime + black + qemp86 + inc86, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.2801  -0.0774  -0.0329   0.0213  17.2152
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.0374053  0.0476873  -0.784   0.4329
## tottime      0.7064354  0.0054793 128.928   <2e-16 ***
## black        0.1402641  0.0690914   2.030   0.0424 *
## qemp86       0.0425101  0.0221607   1.918   0.0552 .
```

```
## inc86       -0.0007928  0.0005335  -1.486   0.1374
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.301 on 2720 degrees of freedom
## Multiple R-squared:  0.8626, Adjusted R-squared:  0.8624
## F-statistic:  4268 on 4 and 2720 DF,  p-value: < 2.2e-16
```

Interpretation of the OLS output

We see a high R-squared. The significant variables for 0.05 significance level are the total time spend in prison and the color. No significance of the other variables is proven.

Problems with the OLS

Some of the variables may be ednogenuos. => Testing this way may not be correct.

IV Regression

The idea is to use instrumental variables in the estimation of the expected severity. Define: endogenuos var: income86, qemp86, tottime exogenuos var: black instruments: durat, nparr, nfarr, narr, ptime86

The regression code is given by:

```
IV_sev1<-ivreg(avgsen~ tottime+ black+ qemp86+ inc86 | black+ durat+ narr86+ nfarr86+ nparr86+ ptime86
```

Check if Instuments are adequate

1. Check if regressors and instruments are correlated

```
i1lm_sev1<- lm(tottime~ black+ durat+ narr86+ nfarr86+ nparr86+ ptime86, data=df)
summary(i1lm_sev1)
```

```
##
## Call:
## lm(formula = tottime ~ black + durat + narr86 + nfarr86 + nparr86 +
##     ptime86, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.254 -0.662 -0.306 -0.281 55.743
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.28080    0.10601   2.649  0.00812 **
## black        1.09852    0.23369   4.701 2.72e-06 ***
## durat        0.02531    0.01845   1.372  0.17013
## narr86       0.38140    0.17599   2.167  0.03031 *
## nfarr86     -0.11213    0.25363  -0.442  0.65845
## nparr86     -0.46308    0.23883  -1.939  0.05261 .
## ptime86      0.65619    0.04338  15.127  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.397 on 2718 degrees of freedom
## Multiple R-squared:  0.09124,    Adjusted R-squared:  0.08923
## F-statistic: 45.48 on 6 and 2718 DF,  p-value: < 2.2e-16
```

```
i2lm_sev1<- lm(qemp86~ black+ durat+ narr86+ nfarr86+ nparr86+ ptime86, data=df)
summary(i2lm_sev1)
```

```
##
## Call:
## lm(formula = qemp86 ~ black + durat + narr86 + nfarr86 + nparr86 +
##     ptime86, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.7583 -0.9233  0.2340  1.0767  4.6960
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.923252   0.031584  92.554  < 2e-16 ***
## black       -0.328207   0.069629  -4.714 2.56e-06 ***
## durat       -0.164969   0.005496 -30.016  < 2e-16 ***
## narr86      -0.157216   0.052437  -2.998  0.00274 **
## nfarr86     -0.159808   0.075569  -2.115  0.03454 *
## nparr86     -0.009608   0.071160  -0.135  0.89260
## ptime86     -0.226935   0.012925 -17.558  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.31 on 2718 degrees of freedom
## Multiple R-squared:  0.3398, Adjusted R-squared:  0.3383
## F-statistic: 233.1 on 6 and 2718 DF,  p-value: < 2.2e-16
```

```
i3lm_sev1<- lm( inc86~ black+ durat+ narr86+ nfarr86+ nparr86+ ptime86, data=df)
summary(i3lm_sev1)
```

```
##
## Call:
## lm(formula = inc86 ~ black + durat + narr86 + nfarr86 + nparr86 +
##     ptime86, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -74.93 -44.37 -16.33  30.17 465.97
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  75.0256     1.4557  51.539  < 2e-16 ***
## black       -14.8285     3.2092  -4.621 4.00e-06 ***
## durat        -4.7147     0.2533 -18.612  < 2e-16 ***
## narr86      -10.3329     2.4168  -4.275 1.97e-05 ***
## nfarr86      -1.7791     3.4829  -0.511    0.610
## nparr86      -2.1158     3.2797  -0.645    0.519
## ptime86      -5.6714     0.5957  -9.520  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 60.38 on 2718 degrees of freedom
## Multiple R-squared:  0.1806, Adjusted R-squared:  0.1788
```

```
## F-statistic: 99.86 on 6 and 2718 DF,  p-value: < 2.2e-16
```

Multiple r-squared > 0 is observed in every regression => first criterion is met.

2. Check if errors and instruments are uncorrelated.

```
resid_sev1<-resid(IV_sev1)
lm_resid_sev1<-lm(resid_sev1~black+ durat+ narr86+ nfarr86+ nparr86+ ptime86, data=df)
summary(lm_resid_sev1)
```

```
##
## Call:
## lm(formula = resid_sev1 ~ black + durat + narr86 + nfarr86 +
##     nparr86 + ptime86, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.7167  -0.1842  -0.0872   0.0257  21.8264
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.003974   0.032882  -0.121    0.904
## black        0.001268   0.072489   0.017    0.986
## durat        0.001107   0.005722   0.193    0.847
## narr86       0.047213   0.054592   0.865    0.387
## nfarr86     -0.017087   0.078673  -0.217    0.828
## nparr86     -0.102002   0.074083  -1.377    0.169
## ptime86     -0.002650   0.013456  -0.197    0.844
##
## Residual standard error: 1.364 on 2718 degrees of freedom
## Multiple R-squared:  0.0008356,  Adjusted R-squared:  -0.00137
## F-statistic: 0.3789 on 6 and 2718 DF,  p-value: 0.8929
```

A really small multiple R-squared is observed. The p-values of variables are considerably higher than 0.05 significance level. What can be done in addition is a test on $n*R^{2}, where R^{2} is the non-centered R^{2}

```
0.0008356*2725
```

```
## [1] 2.27701
```

Which is smaller than the Chi-square value on 2 df and 0.05 significanse level=> also the second criterion is met.

#####Summary of the IV-Model

```
summary(IV_sev1)
```

```
##
## Call:
## ivreg(formula = avgsen ~ tottime + black + qemp86 + inc86 | black +
##     durat + narr86 + nfarr86 + nparr86 + ptime86, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -11.72068  -0.17947  -0.09283   0.02787  21.82055
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)  0.122938   0.132308   0.929  0.35288
## tottime      0.627074   0.026174  23.958  < 2e-16 ***
## black        0.258763   0.090411   2.862  0.00424 **
## qemp86      -0.099968   0.185116  -0.540  0.58922
## inc86        0.003139   0.006011   0.522  0.60157
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.364 on 2720 degrees of freedom
## Multiple R-Squared: 0.8491,  Adjusted R-squared: 0.8488
## Wald test:   309 on 4 and 2720 DF,  p-value: < 2.2e-16
```

Here a high R-squared is observed. Tottime and Black are the only significant variables.

##############Georgi#################### ##############################

Simple OLS Regression, LPM

OLS estimation

```
##
## Call:
## lm(formula = narr86 ~ pcnv + avgsen + tottime + ptime86 + qemp86 +
##     inc86 + durat + black + hispan + born60 + pcnvsq + pt86sq +
##     inc86sq, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5542 -0.4622 -0.2097  0.2374 11.3955
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.618e-01  4.481e-02  12.537  < 2e-16 ***
## pcnv         5.710e-01  1.544e-01   3.697 0.000222 ***
## avgsen      -1.708e-02  1.205e-02  -1.418 0.156417
## tottime      1.203e-02  9.277e-03   1.297 0.194806
## ptime86      2.936e-01  4.432e-02   6.624 4.19e-11 ***
## qemp86      -2.706e-02  1.840e-02  -1.471 0.141512
## inc86       -3.348e-03  8.048e-04  -4.160 3.28e-05 ***
## durat       -7.652e-03  3.962e-03  -1.931 0.053535 .
## black        2.936e-01  4.481e-02   6.551 6.80e-11 ***
## hispan       1.616e-01  3.944e-02   4.098 4.29e-05 ***
## born60      -3.767e-02  3.278e-02  -1.149 0.250623
## pcnvsq      -7.488e-01  1.563e-01  -4.792 1.74e-06 ***
## pt86sq      -3.044e-02  3.879e-03  -7.846 6.12e-15 ***
## inc86sq      7.148e-06  2.555e-06   2.798 0.005178 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8146 on 2711 degrees of freedom
## Multiple R-squared:  0.1051, Adjusted R-squared:  0.1008
## F-statistic:  24.5 on 13 and 2711 DF,  p-value: < 2.2e-16
```

Model:

$narr86 = \beta_0 + \beta_1\ pcnv + \beta_2\ ptime86 + \beta_3\ inc86 + \beta_4\ black + \beta_5\ hispan + \beta_6\ pcnvsq + \beta_7\ pt86sq + \beta_8\ inc86sq$

```
##
## Call:
## lm(formula = narr86 ~ pcnv + ptime86 + inc86 + black + hispan +
##     pcnvsq + pt86sq + inc86sq, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5498 -0.4692 -0.2159  0.2309 11.4326
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.896e-01  3.227e-02  15.173  < 2e-16 ***
## pcnv         5.500e-01  1.533e-01   3.587  0.00034 ***
## ptime86      2.880e-01  4.388e-02   6.563 6.30e-11 ***
## inc86       -3.906e-03  5.257e-04  -7.430 1.45e-13 ***
## black        2.908e-01  4.464e-02   6.514 8.71e-11 ***
## hispan       1.623e-01  3.938e-02   4.120 3.89e-05 ***
## pcnvsq      -7.286e-01  1.552e-01  -4.695 2.80e-06 ***
## pt86sq      -2.946e-02  3.850e-03  -7.652 2.72e-14 ***
## inc86sq      8.377e-06  2.096e-06   3.996 6.60e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.815 on 2716 degrees of freedom
## Multiple R-squared:  0.1026, Adjusted R-squared:  0.09991
## F-statistic:  38.8 on 8 and 2716 DF,  p-value: < 2.2e-16
```

IV Regression

Model:

Generalized IV estimation

2SLS estimation

LOGIT model

creating a binary variable arr86, when a person gets arrested at least once.

```
df$arr86 <- ifelse(df$narr86>0 ,1 ,0)
```

```
log_all <- glm(arr86 ~ pcnv + avgsen + tottime + ptime86 + qemp86 + inc86 + durat + black + hispan + bo

summary(log_all)
```

```
##
## Call:
## glm(formula = arr86 ~ pcnv + avgsen + tottime + ptime86 + qemp86 +
##     inc86 + durat + black + hispan + born60 + pcnvsq + pt86sq +
##     inc86sq, family = binomial(link = "logit"), data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1656  -0.8658  -0.5644   1.1201   2.6271
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept) -7.302e-01  1.225e-01  -5.960 2.53e-09 ***
## pcnv         4.390e-01  4.348e-01   1.010 0.312619
## avgsen       2.614e-02  4.384e-02   0.596 0.550956
## tottime     -3.245e-02  3.562e-02  -0.911 0.362387
## ptime86      1.263e+00  2.523e-01   5.007 5.52e-07 ***
## qemp86       1.373e-01  5.144e-02   2.669 0.007607 **
## inc86       -1.448e-02  2.471e-03  -5.860 4.63e-09 ***
## durat        1.235e-02  1.039e-02   1.189 0.234550
## black        7.322e-01  1.209e-01   6.058 1.38e-09 ***
## hispan       4.386e-01  1.129e-01   3.886 0.000102 ***
## born60      -1.587e-02  9.635e-02  -0.165 0.869192
## pcnvsq      -1.552e+00  4.618e-01  -3.361 0.000776 ***
## pt86sq      -1.742e-01  3.911e-02  -4.453 8.48e-06 ***
## inc86sq      2.468e-05  8.186e-06   3.015 0.002570 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3216.4  on 2724  degrees of freedom
## Residual deviance: 2871.9  on 2711  degrees of freedom
## AIC: 2899.9
##
## Number of Fisher Scoring iterations: 8
```

```
log <- glm(arr86 ~   ptime86 + qemp86 + inc86  + black + hispan + pcnvsq + pt86sq + inc86sq , data = df

summary(log)
```

```
##
## Call:
## glm(formula = arr86 ~ ptime86 + qemp86 + inc86 + black + hispan +
##     pcnvsq + pt86sq + inc86sq, family = binomial(link = "logit"),
##     data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1653  -0.8654  -0.5673   1.1359   2.6267
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -6.312e-01  9.372e-02  -6.735 1.64e-11 ***
## ptime86      1.251e+00  2.467e-01   5.070 3.97e-07 ***
## qemp86       1.175e-01  4.857e-02   2.420   0.0155 *
## inc86       -1.458e-02  2.459e-03  -5.929 3.05e-09 ***
## black        7.297e-01  1.202e-01   6.073 1.26e-09 ***
## hispan       4.471e-01  1.116e-01   4.008 6.13e-05 ***
## pcnvsq      -1.114e+00  1.379e-01  -8.079 6.55e-16 ***
## pt86sq      -1.733e-01  3.847e-02  -4.504 6.67e-06 ***
## inc86sq      2.480e-05  8.170e-06   3.036   0.0024 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
##     Null deviance: 3216.4  on 2724  degrees of freedom
## Residual deviance: 2875.7  on 2716  degrees of freedom
## AIC: 2893.7
##
## Number of Fisher Scoring iterations: 8
```

For comparison

```
prob <- glm(arr86 ~   ptime86 + qemp86 + inc86  + black + hispan + pcnvsq + pt86sq + inc86sq , data = d

summary(prob)
```

```
##
## Call:
## glm(formula = arr86 ~ ptime86 + qemp86 + inc86 + black + hispan +
##     pcnvsq + pt86sq + inc86sq, family = binomial(link = "probit"),
##     data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1724  -0.8682  -0.5697   1.1467   2.7138
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.917e-01  5.648e-02  -6.936 4.04e-12 ***
## ptime86      7.387e-01  1.400e-01   5.278 1.31e-07 ***
## qemp86       6.771e-02  2.898e-02   2.337  0.01944 *
## inc86       -8.503e-03  1.417e-03  -6.001 1.96e-09 ***
## black        4.373e-01  7.299e-02   5.992 2.08e-09 ***
## hispan       2.615e-01  6.643e-02   3.936 8.28e-05 ***
## pcnvsq      -6.503e-01  7.687e-02  -8.461  < 2e-16 ***
## pt86sq      -1.021e-01  2.183e-02  -4.676 2.93e-06 ***
## inc86sq      1.520e-05  4.623e-06   3.287  0.00101 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3216.4  on 2724  degrees of freedom
## Residual deviance: 2876.2  on 2716  degrees of freedom
## AIC: 2894.2
##
## Number of Fisher Scoring iterations: 8
```

Calculation of MC Faddens pseudo R^2

```
1-(prob$deviance/prob$null.deviance)
```

```
## [1] 0.105748
```

parameter logit probit check

prob/log slide 21 ... factor 1.6

##Average marginal effect

```
# for logit
fav <- mean(dnorm(predict(log,type="link")))
fav*coef(log)
```

```
##   (Intercept)       ptime86        qemp86         inc86         black
## -1.391845e-01  2.758107e-01  2.591507e-02 -3.214709e-03  1.608863e-01
##        hispan        pcnvsq        pt86sq       inc86sq
##  9.857880e-02 -2.456187e-01 -3.820432e-02  5.468947e-06
```

Predicition of cutoff

```
tab <- table(true= df$arr86, pred= round(fitted(log)))
tab
```

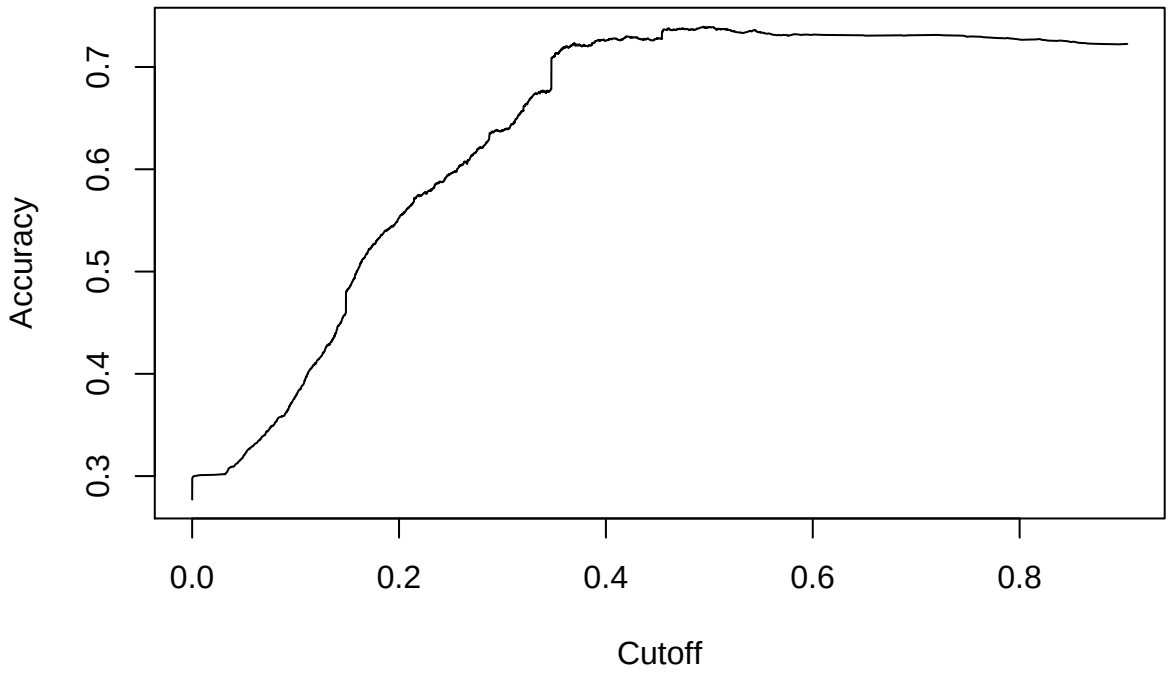
```
##     pred
## true    0    1
##    0 1883   87
##    1  625  130
```

```
tabp <-round(100* c(tab[1,1]+tab[2,2], tab[2,1]+ tab[1,2]/sum(tab)), digits = 2)
tabp
```

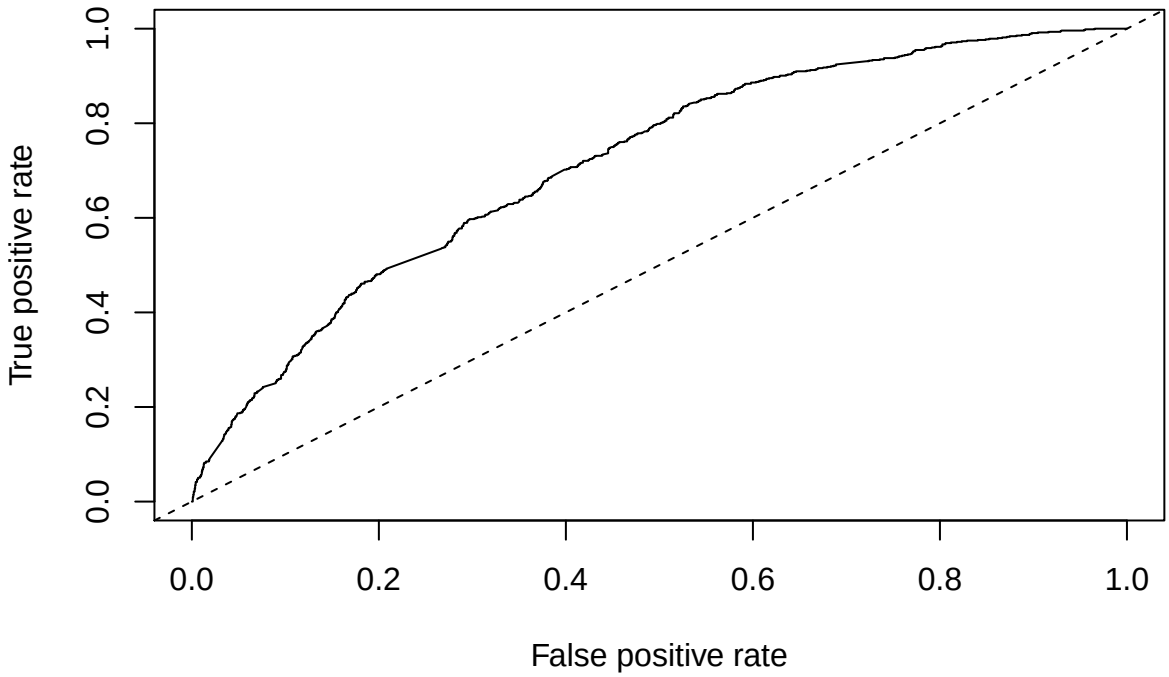
```
## [1] 201300.00  62503.19
```

ROC

```
pred <- prediction(fitted(log),df$arr86)
plot(performance(pred, "acc"))
```




```
plot(performance(pred,"tpr", "fpr"))
abline(0,1,lty=2)
```



Ordered Logit Model

```
library("oglmx")
```

```
## Loading required package: maxLik
```

```
## Loading required package: miscTools
```

```
##
## Please cite the 'maxLik' package as:
## Henningsen, Arne and Toomet, Ott (2011). maxLik: A package for maximum likelihood estimation in R. C
##
## If you have questions, suggestions, or comments regarding the 'maxLik' package, please use a forum o
## https://r-forge.r-project.org/projects/maxlik/
```

```
results.oprob<-oglmx(narr86 ~  ptime86 + qemp86 + inc86  + black + hispan + pcnvsq + pt86sq + inc86sq,
                     constantMEAN = FALSE, constantSD = FALSE,
                     delta=0,threshparam = NULL)
summary(results.oprob)
```

```
## Ordered Probit Regression
## Log-Likelihood: -2080.282
## No. Iterations: 10
## McFadden's R2: 0.8733637
## AIC: 4196.564
##            Estimate  Std. error t value  Pr(>|t|)
## ptime86  7.1443e-01  1.1098e-01  6.4374 1.216e-10 ***
## qemp86   3.5043e-02  2.7477e-02  1.2753  0.202185
## inc86   -8.2362e-03  1.3779e-03 -5.9773 2.268e-09 ***
## black    4.4770e-01  6.7263e-02  6.6560 2.814e-11 ***
## hispan   3.0238e-01  6.2454e-02  4.8416 1.288e-06 ***
## pcnvsq  -5.5297e-01  7.3526e-02 -7.5208 5.444e-14 ***
## pt86sq  -1.0195e-01  1.8300e-02 -5.5710 2.532e-08 ***
```

```
## inc86sq  1.5006e-05  4.5761e-06  3.2792  0.001041 **
## ----- Threshold Parameters -----
##                     Estimate Std. error t value  Pr(>|t|)
## Threshold (0->1)   0.361356   0.054279  6.6575 2.786e-11 ***
## Threshold (1->2)   1.338313   0.060388 22.1620 < 2.2e-16 ***
## Threshold (2->3)   1.851261   0.070803 26.1468 < 2.2e-16 ***
## Threshold (3->4)   2.227259   0.086089 25.8717 < 2.2e-16 ***
## Threshold (4->5)   2.416535   0.097762 24.7185 < 2.2e-16 ***
## Threshold (5->6)   2.774494   0.131885 21.0373 < 2.2e-16 ***
## Threshold (6->7)   3.001412   0.167156 17.9557 < 2.2e-16 ***
## Threshold (7->9)   3.092331   0.186322 16.5967 < 2.2e-16 ***
## Threshold (9->10)  3.218277   0.218835 14.7064 < 2.2e-16 ***
## Threshold (10->12) 3.416592   0.285862 11.9519 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
"marginal effects"
```

```
## [1] "marginal effects"
```

```
margins.oglmx(results.oprob,ascontinuous = TRUE)
```

```
## Marginal Effects on Pr(Outcome==0)
##           Marg. Eff  Std. error t value  Pr(>|t|)
## ptime86 -2.1244e-01  2.8908e-02 -7.3490 1.998e-13 ***
## qemp86  -1.0420e-02  8.1619e-03 -1.2767  0.201705
## inc86    2.4491e-03  4.1310e-04  5.9285 3.056e-09 ***
## black   -1.3313e-01  2.0307e-02 -6.5558 5.533e-11 ***
## hispan  -8.9915e-02  1.8699e-02 -4.8085 1.521e-06 ***
## pcnvsq   1.6443e-01  2.1998e-02  7.4750 7.721e-14 ***
## pt86sq   3.0316e-02  4.7373e-03  6.3994 1.560e-10 ***
## inc86sq -4.4621e-06  1.3655e-06 -3.2678  0.001084 **
## ------------------------------------
## Marginal Effects on Pr(Outcome==1)
##           Marg. Eff  Std. error t value  Pr(>|t|)
## ptime86  1.5011e-01  2.2600e-02  6.6420 3.094e-11 ***
## qemp86   7.3630e-03  5.7731e-03  1.2754  0.202164
## inc86   -1.7305e-03  2.9564e-04 -5.8535 4.813e-09 ***
## black    9.4068e-02  1.4567e-02  6.4575 1.064e-10 ***
## hispan   6.3533e-02  1.3330e-02  4.7663 1.876e-06 ***
## pcnvsq  -1.1619e-01  1.5868e-02 -7.3219 2.445e-13 ***
## pt86sq  -2.1421e-02  3.6933e-03 -5.7999 6.634e-09 ***
## inc86sq  3.1529e-06  9.6822e-07  3.2564  0.001128 **
## ------------------------------------
## Marginal Effects on Pr(Outcome==2)
##           Marg. Eff  Std. error t value  Pr(>|t|)
## ptime86  3.9988e-02  5.8018e-03  6.8923 5.491e-12 ***
## qemp86   1.9614e-03  1.5431e-03  1.2711  0.203690
## inc86   -4.6100e-04  8.7744e-05 -5.2538 1.490e-07 ***
## black    2.5059e-02  4.4312e-03  5.6550 1.558e-08 ***
## hispan   1.6925e-02  3.8379e-03  4.4098 1.034e-05 ***
## pcnvsq  -3.0951e-02  4.9188e-03 -6.2924 3.127e-10 ***
## pt86sq  -5.7063e-03  9.0245e-04 -6.3231 2.563e-10 ***
## inc86sq  8.3990e-07  2.6776e-07  3.1368  0.001708 **
## ------------------------------------
```

```
## Marginal Effects on Pr(Outcome==3)
##           Marg. Eff  Std. error t value  Pr(>|t|)
## ptime86  1.3431e-02  2.4802e-03  5.4154 6.113e-08 ***
## qemp86   6.5881e-04  5.2618e-04  1.2521 0.2105464
## inc86   -1.5484e-04  3.6121e-05 -4.2868 1.813e-05 ***
## black    8.4168e-03  1.8693e-03  4.5026 6.714e-06 ***
## hispan   5.6847e-03  1.4941e-03  3.8049 0.0001419 ***
## pcnvsq  -1.0396e-02  2.1838e-03 -4.7605 1.931e-06 ***
## pt86sq  -1.9167e-03  3.6574e-04 -5.2406 1.601e-07 ***
## inc86sq  2.8211e-07  9.7644e-08  2.8892 0.0038625 **
## ---------------------------------------
## Marginal Effects on Pr(Outcome==4)
##           Marg. Eff  Std. error t value Pr(>|t|)
## ptime86  3.5933e-03  1.1010e-03  3.2637 0.001100 **
## qemp86   1.7625e-04  1.4786e-04  1.1920 0.233255
## inc86   -4.1425e-05  1.4253e-05 -2.9064 0.003656 **
## black    2.2518e-03  7.5692e-04  2.9749 0.002931 **
## hispan   1.5208e-03  5.5202e-04  2.7550 0.005868 **
## pcnvsq  -2.7812e-03  9.1744e-04 -3.0315 0.002433 **
## pt86sq  -5.1277e-04  1.5825e-04 -3.2403 0.001194 **
## inc86sq  7.5474e-08  3.2351e-08  2.3330 0.019651 *
## ---------------------------------------
## Marginal Effects on Pr(Outcome==5)
##           Marg. Eff  Std. error t value  Pr(>|t|)
## ptime86  3.5017e-03  1.0569e-03  3.3133 0.0009222 ***
## qemp86   1.7176e-04  1.4465e-04  1.1875 0.2350452
## inc86   -4.0369e-05  1.3993e-05 -2.8851 0.0039132 **
## black    2.1944e-03  7.4258e-04  2.9551 0.0031257 **
## hispan   1.4821e-03  5.3946e-04  2.7473 0.0060081 **
## pcnvsq  -2.7104e-03  9.0105e-04 -3.0080 0.0026295 **
## pt86sq  -4.9970e-04  1.5130e-04 -3.3028 0.0009574 ***
## inc86sq  7.3551e-08  3.1676e-08  2.3220 0.0202336 *
## ---------------------------------------
## Marginal Effects on Pr(Outcome==6)
##           Marg. Eff  Std. error t value Pr(>|t|)
## ptime86  9.5621e-04  5.0382e-04  1.8979  0.05771 .
## qemp86   4.6903e-05  4.4592e-05  1.0518  0.29289
## inc86   -1.1024e-05  6.1462e-06 -1.7936  0.07288 .
## black    5.9922e-04  3.3051e-04  1.8130  0.06983 .
## hispan   4.0471e-04  2.3006e-04  1.7592  0.07855 .
## pcnvsq  -7.4012e-04  4.0640e-04 -1.8211  0.06858 .
## pt86sq  -1.3645e-04  7.1834e-05 -1.8996  0.05749 .
## inc86sq  2.0084e-08  1.2316e-08  1.6307  0.10295
## ---------------------------------------
## Marginal Effects on Pr(Outcome==7)
##           Marg. Eff  Std. error t value Pr(>|t|)
## ptime86  2.3179e-04  2.3609e-04  0.9818   0.3262
## qemp86   1.1370e-05  1.4692e-05  0.7739   0.4390
## inc86   -2.6722e-06  2.7667e-06 -0.9658   0.3341
## black    1.4526e-04  1.4983e-04  0.9695   0.3323
## hispan   9.8106e-05  1.0212e-04  0.9607   0.3367
## pcnvsq  -1.7941e-04  1.8500e-04 -0.9698   0.3322
## pt86sq  -3.3078e-05  3.3673e-05 -0.9823   0.3259
## inc86sq  4.8686e-09  5.1918e-09  0.9378   0.3484
```

```
## ------------------------------------
## Marginal Effects on Pr(Outcome==9)
##           Marg. Eff  Std. error t value Pr(>|t|)
## ptime86  2.2718e-04  2.3201e-04  0.9792   0.3275
## qemp86   1.1143e-05  1.4437e-05  0.7719   0.4402
## inc86   -2.6191e-06  2.7195e-06 -0.9631   0.3355
## black    1.4237e-04  1.4722e-04  0.9670   0.3335
## hispan   9.6154e-05  1.0033e-04  0.9584   0.3379
## pcnvsq  -1.7584e-04  1.8191e-04 -0.9666   0.3337
## pt86sq  -3.2419e-05  3.3084e-05 -0.9799   0.3271
## inc86sq  4.7718e-09  5.1022e-09  0.9352   0.3497
## ------------------------------------
## Marginal Effects on Pr(Outcome==10)
##           Marg. Eff  Std. error t value Pr(>|t|)
## ptime86  2.0960e-04  2.1802e-04  0.9614   0.3364
## qemp86   1.0281e-05  1.3490e-05  0.7621   0.4460
## inc86   -2.4163e-06  2.5551e-06 -0.9457   0.3443
## black    1.3135e-04  1.3821e-04  0.9503   0.3419
## hispan   8.8711e-05  9.4251e-05  0.9412   0.3466
## pcnvsq  -1.6223e-04  1.7092e-04 -0.9492   0.3425
## pt86sq  -2.9910e-05  3.1081e-05 -0.9623   0.3359
## inc86sq  4.4024e-09  4.7887e-09  0.9193   0.3579
## ------------------------------------
## Marginal Effects on Pr(Outcome==12)
##           Marg. Eff  Std. error t value Pr(>|t|)
## ptime86  1.9188e-04  2.0799e-04  0.9225   0.3563
## qemp86   9.4119e-06  1.2707e-05  0.7407   0.4589
## inc86   -2.2121e-06  2.4364e-06 -0.9079   0.3639
## black    1.2024e-04  1.3188e-04  0.9118   0.3619
## hispan   8.1213e-05  8.9931e-05  0.9031   0.3665
## pcnvsq  -1.4852e-04  1.6298e-04 -0.9113   0.3622
## pt86sq  -2.7382e-05  2.9638e-05 -0.9239   0.3556
## inc86sq  4.0303e-09  4.5559e-09  0.8846   0.3764
```

##WORK IN PROGRESS:Alternative model with fixed thresholds (restrictions)

```
#results.oprob1<-oglmx(y ~ x1 + x2, ~ x1 + x2, data=df,
#                       constantMEAN = FALSE, constantSD = FALSE)
"Alternative model with fixed thresholds"
```

```
## [1] "Alternative model with fixed thresholds"
```

```
#results.oprob1alt<-oglmx(y ~ x1 + x2, ~ x1 + x2, data=df,
#                       constantMEAN = TRUE, constantSD = TRUE,
#                       threshparam=c(-0.5,NA,1.5))
#summary(results.oprob1)
#summary(results.oprob1alt)
```

##WORK IN PROGRESS:Some Testing

```
library("lmtest")

#lrtest(results.oprob,results.oprobalt)
```

Truncated model, let's see