Econometrics in-class Assignment

1. As in standard criminometric studies, consider the following model to explain the crime rate (the number of crimes per person, crmrte) in terms of the estimated probability of arrest (prbarr), the estimated probability of convinction given an arrest (prbconv), the average sentence length served (avgsen) and the number of police officers per capita (polpc):

```
Call:
lm(formula = log(crmrte) ~ log(prbarr) + log(prbconv) + log(avgsen) +
    log(polpc), data = crime)

Residuals:
     Min       1Q   Median       3Q      Max
-1.85569 -0.18673  0.03436  0.23903  1.24497

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -2.43782    0.23198 -10.509   <2e-16 ***
log(prbarr)  -0.72629    0.03703 -19.613   <2e-16 ***
log(prbconv) -0.56356    0.02621 -21.502   <2e-16 ***
log(avgsen)  -0.06575    0.05587  -1.177     0.24
log(polpc)    0.36141    0.03024  11.951   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3825 on 625 degrees of freedom
Multiple R-squared:  0.5569,    Adjusted R-squared:  0.5541
F-statistic: 196.4 on 4 and 625 DF,  p-value: < 2.2e-16
```

We get the following results:

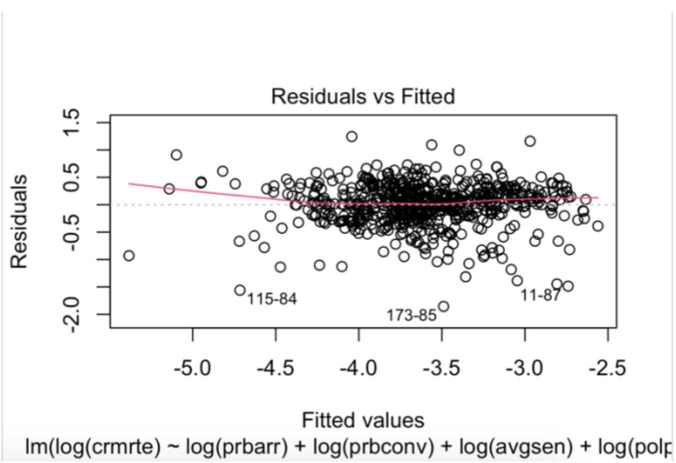
Our results show that for a 1 percent increase in the estimated probability of arrest, holding all other factors fixed, decreases the crime rate by approximately 0.7%.

We also observe a negative coefficient for log of prbconv: a 1 % increase in the estimated probability of conviction decreased the crime rate by approximately 0.5%, ceteris paribus.

Similarly, we observe a negative coefficient for avgsen: a: a 1 % increase in the average length of sentence served, decreases the crime rate by approximately 0.06%, ceteris paribus.

We can see our results are very statistically significant except for the average sentence served, which is not significant at all.
Additionally, our R square value is 0.54 meaning 54% of the variability is explained by the model.

Looking at the residuals vs. fitted plot for model 1, we observe that as x increases, the residuals seem more dispersed. This could possibly be due to heteroskedasticity in our model



2. Compute the Breusch-Pagan test for heteroskedasticity. What do you conclude? Report the heteroskedasticity-robust standard errors.

```
        studentized Breusch-Pagan test

data:  model1
BP = 114.17, df = 4, p-value < 2.2e-16
```

Since P-value is less than 0.01, we can reject the null hypothesis meaning we have sufficient evidence to say that heteroscedasticity is present in the regression model.
Since we know that the model1 suffers from heteroskedasticity, we want to obtain heteroskedasticity robust standard errors and their corresponding t values using the coeftest:

```
t test of coefficients:

                 Estimate  Std. Error  t value  Pr(>|t|)    
(Intercept)    0.09911172  0.01503219   6.5933 9.160e-11 ***
log(prbarr)   -0.02152591  0.00158387 -13.5907 < 2.2e-16 ***
log(prbconv)  -0.01575190  0.00110149 -14.3006 < 2.2e-16 ***
log(avgsen)    0.00060439  0.00215549   0.2804    0.7793    
log(polpc)     0.01651149  0.00204764   8.0637 3.785e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

3. Why could we suspect that the variable polpc is endogenous? Consider density (people per sq. mile) and taxpc (tax revenue per capita) as potential IV for polpc and run the following reduced form equation:

We could expect polpc to be endogenous because the probability of arrest could depend on the number of police officers working at that time. Additionally, the presence of police affects crime rates and or rates of arrest which in turn affect the number of police (if crime rate increases, so does number of police, usually).

We see that our results are all statistically significant except for average sentence length served which is insignificant (same as above).

```
Call:
lm(formula = log(polpc) ~ log(prbarr) + log(prbconv) + log(avgsen) + 
    log(density) + log(taxpc), data = crime)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.48220 -0.24160 -0.01247  0.17491  2.70304 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -7.35732    0.26522 -27.740  < 2e-16 ***
log(prbarr)   0.26749    0.05133   5.211 2.55e-07 ***
log(prbconv)  0.34675    0.03540   9.795  < 2e-16 ***
log(avgsen)   0.06102    0.06928   0.881    0.379    
log(density)  0.19186    0.02999   6.398 3.09e-10 ***
log(taxpc)    0.39312    0.06157   6.385 3.34e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4743 on 624 degrees of freedom
Multiple R-squared:  0.1955,    Adjusted R-squared:  0.189 
F-statistic: 30.32 on 5 and 624 DF,  p-value: < 2.2e-16
```

For example, a 1% increase in the probability of arrest increases the number of police officers per capita by approximately 0.26%.

A 1% increase in the people per square mile (density) increases the police officers per capita by approximately 0.2%, ceteris paribus.
A 1% increase in the tax revenue per capita increases the police officers per capita by approximately 0.4%, ceteris paribus.

We can see that while the results are statistically significant, the coefficient remains low.

4. Use density and taxpc as IV to carry out a 2SLS estimation of model (1). Interpret the results you obtained.

```
Residuals:
     Min       1Q   Median       3Q      Max 
-2.73994 -0.19904  0.05232  0.28817  1.39148 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)         1.92740    0.89063   2.164   0.0308 *  
log(crime$prbarr)  -0.79806    0.08848  -9.019  < 2e-16 ***
log(crime$prbconv) -0.73055    0.05488 -13.312  < 2e-16 ***
log(crime$avgsen)  -0.12532    0.10630  -1.179   0.2388    
log(crime$polpc)    1.04611    0.12930   8.090 3.11e-15 ***

Diagnostic tests:
                 df1 df2 statistic  p-value    
Weak instruments   2 624     43.10  < 2e-16 ***
Wu-Hausman         1 624     50.40 3.42e-12 ***
Sargan             1  NA     19.44 1.04e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5161 on 625 degrees of freedom
Multiple R-Squared: 0.1935,	Adjusted R-squared: 0.1883 
Wald test: 72.59 on 4 and 625 DF,  p-value: < 2.2e-16
```

we can see that our instruments are not weak and all results for IV’s are significant. This means that our IV’s are good in order to combat the endogeneity bias.
As we have more than one IV, looking at the Sargan test allows us to conclude that our IV model is satisfactory.