Exercise 6

Problem 1

The file *stockton96.gdt* contains 940 observations on home sales in Stockton, CA in 1996.

a) Use least squares to estimate a linear equation that relates house price *PRICE* to the size of the house in square feet *SQFT* and the age of the house in years *AGE*. Interpret all the estimates.

ols price const age sqft

```
Model 1: OLS, using observations 1-940
Dependent variable: price

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       5193.15      3586.64        1.448    0.1480
  age         -217.843       35.0976     -6.207    8.11e-010 ***
  sqft          68.3907       2.16868    31.54     2.39e-149 ***

Mean dependent var   97937.83   S.D. dependent var   34179.37
Sum squared resid    4.76e+11   S.E. of regression   22539.63
R-squared            0.566050   Adjusted R-squared   0.565124
F(2, 937)            611.1178   P-value(F)           1.4e-170
Log-likelihood      -10753.95   Akaike criterion     21513.90
Schwarz criterion    21528.43   Hannan-Quinn         21519.44
```

b) Suppose that you own two houses. One has 1400 square feet; the other has 1800 square feet. Both are 20 years old. What price do you estimate you will get for each house?

$$\widehat{p_1} = 5193 + 20 * (-217) + 68.39 * 1400$$
$$\widehat{p_2} = 5193 + 20 * (-217) + 68.39 * 1800$$

c) Test the hypothesis that the size and the age of the house are important determinants of its price (separately as well as jointly). **Both have three stars. Also jointly significant according to above output**

d) Using the Breusch-Pagan test for heteroscedasticity, test whether the model satisfies the homoscedasticity assumption by using the command for the BP test in Gretl.

series yhat=$yhat
genr resid=price-yhat
modtest --breusch-pagan

e) Use the White test to test for heteroskedasticity.

modtest --white

f) What do you conclude regarding the heteroskedasticity? Does your conclusion depend on the choosing a specific test? Discuss also drawbacks of the BP and White tests.

There is heteroskedasticity

A weakness of the BP test is that it assumes the heteroskedasticity is a linear function of the independent variables. Failing to find evidence of heteroskedasticity with the BP doesn't rule out a nonlinear relationship between the independent variable(s) and the error variance.

The weakness of white test is that if you have many variables, the number of possible interactions plus the squared variables plus the original variables can be quite high.

g) Test the hypothesis that the size and the age of the house are important determinants of its price (separately as well as jointly). Hint: choose appropriate standard errors. Does your conclusion differ from part (c)?

ols price const age sqft –robust

compare the robust and non-robust standard errors and parameters. You can see that the parameters did not change, while standard errors increased. Still, conclusions have not changed, based on the F-statistic

```
? ols price const sqft age --robust

Model 10: OLS, using observations 1-940
Dependent variable: price
Heteroskedasticity-robust standard errors, variant HC1

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       5193.15      3648.56        1.423    0.1550
  sqft          68.3907       2.46807    27.71     6.35e-124 ***
  age         -217.843       36.3142     -5.999    2.84e-09  ***

Mean dependent var   97937.83   S.D. dependent var   34179.37
Sum squared resid    4.76e+11   S.E. of regression   22539.63
R-squared            0.566050   Adjusted R-squared   0.565124
F(2, 937)            476.5571   P-value(F)           1.7e-143
Log-likelihood      -10753.95   Akaike criterion     21513.90
Schwarz criterion    21528.43   Hannan-Quinn         21519.44

? ols price const sqft age

Model 11: OLS, using observations 1-940
Dependent variable: price

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       5193.15      3586.64        1.448    0.1480
  sqft          68.3907       2.16868    31.54     2.39e-149 ***
  age         -217.843       35.0976     -6.207    8.11e-010 ***

Mean dependent var   97937.83   S.D. dependent var   34179.37
Sum squared resid    4.76e+11   S.E. of regression   22539.63
R-squared            0.566050   Adjusted R-squared   0.565124
F(2, 937)            611.1178   P-value(F)           1.4e-170
Log-likelihood      -10753.95   Akaike criterion     21513.90
Schwarz criterion    21528.43   Hannan-Quinn         21519.44
```

Problem 2

Using the data in *cps4_small.gdt* estimate the following wage equation with least squares and heteroskedasticity-robust standard errors:

$$\ln(WAGE) = \beta_1 + \beta_2 EDUC + \beta_3 EXPER + \beta_4 EXPER^2 + \beta_5 (EXPERxEDUC) + e$$

(a) Report the results.

genr exper2=exper^2
genr experedu=exper*educ
genr lnwage=ln(wage)
ols lnwage educ exper exper2 experedu const --robust

```
? ols lnwage educ exper exper2 experedu const --robust

Model 4: OLS, using observations 1-1000
Dependent variable: lnwage
Heteroskedasticity-robust standard errors, variant HC1

             coefficient    std. error    t-ratio    p-value
  -----------------------------------------------------------
  const       0.529677      0.252825       2.095     0.0364    **
  educ        0.127195      0.0169597      7.500     1.41e-013 ***
  exper       0.0629807     0.0113775      5.536     3.97e-08  ***
  exper2     -0.000713939   9.20134e-05   -7.759     2.11e-014 ***
  experedu   -0.00132239    0.000636794   -2.077     0.0381    **

Mean dependent var   2.856988   S.D. dependent var   0.580619
Sum squared resid    254.4216   S.E. of regression   0.505668
R-squared            0.244548   Adjusted R-squared   0.241511
F(4, 995)            85.06746   P-value(F)           3.57e-62
Log-likelihood      -734.5572   Akaike criterion     1479.114
Schwarz criterion    1503.653   Hannan-Quinn         1488.441
```

(b) Add MARRIED to the equation and re-estimate. Holding education and experience constant, do married workers get higher wages? Using a 5% significance level, test a null hypothesis that wages of married workers are less than or equal to those of unmarried workers against the alternative that wages of married workers are higher.

```
? ols lnwage educ exper exper2 experedu married const --robust

Model 5: OLS, using observations 1-1000
Dependent variable: lnwage
Heteroskedasticity-robust standard errors, variant HC1

             coefficient    std. error    t-ratio    p-value
  -----------------------------------------------------------
  const       0.541061      0.254209       2.128     0.0335    **
  educ        0.126120      0.0170564      7.394     3.02e-013 ***
  exper       0.0613731     0.0115877      5.296     1.45e-07  ***
  exper2     -0.000693346   9.55671e-05   -7.255     8.07e-013 ***
  experedu   -0.00130912    0.000638420   -2.051     0.0406    **
  married     0.0402895     0.0339231      1.188     0.2352

Mean dependent var   2.856988   S.D. dependent var   0.580619
Sum squared resid    254.0582   S.E. of regression   0.505561
R-squared            0.245627   Adjusted R-squared   0.241833
F(5, 994)            69.11228   P-value(F)           4.41e-62
Log-likelihood      -733.8426   Akaike criterion     1479.685
Schwarz criterion    1509.132   Hannan-Quinn         1490.877
```

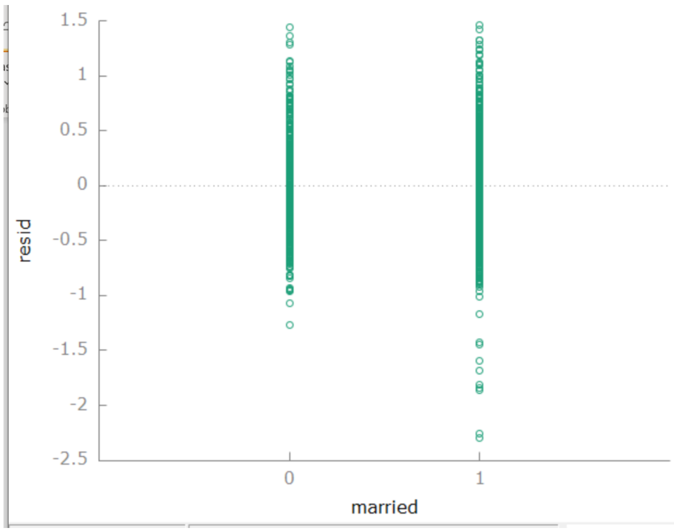
The null and alternative hypotheses for testing whether married workers get higher wages are given by

$$\boldsymbol{H_0: \beta_6 \leq 0}$$
$$\boldsymbol{H_1: \beta_6 > 0}$$

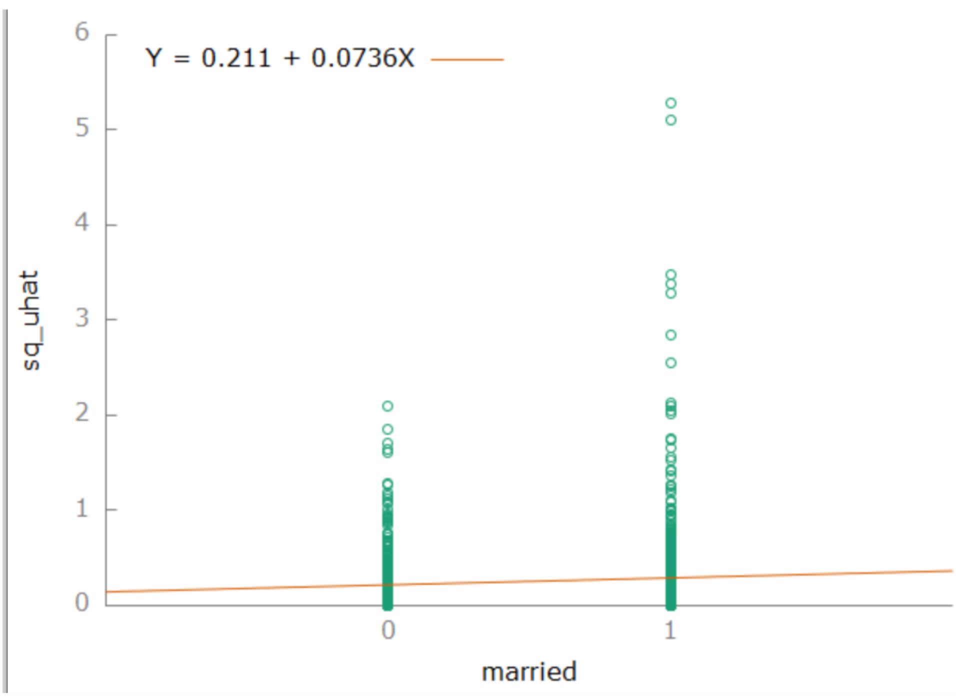
The test value is: 1.188, the critical value at the 5% level of significance is 1.646. Since the test value is less than the critical value, we do not reject the null hypothesis at the 5% level. We conclude that there is insufficient evidence to show that wages of married workers are greater than those of unmarried workers.

(c) Plot the residuals from part (a) against the two values of MARRIED. Is there evidence of heteroskedasticity?

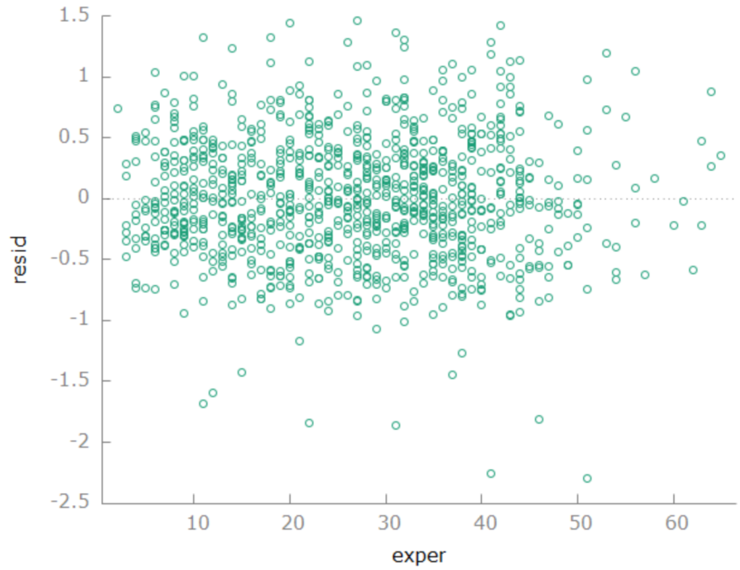
series uhat=$uhat
gnuplot uhat married

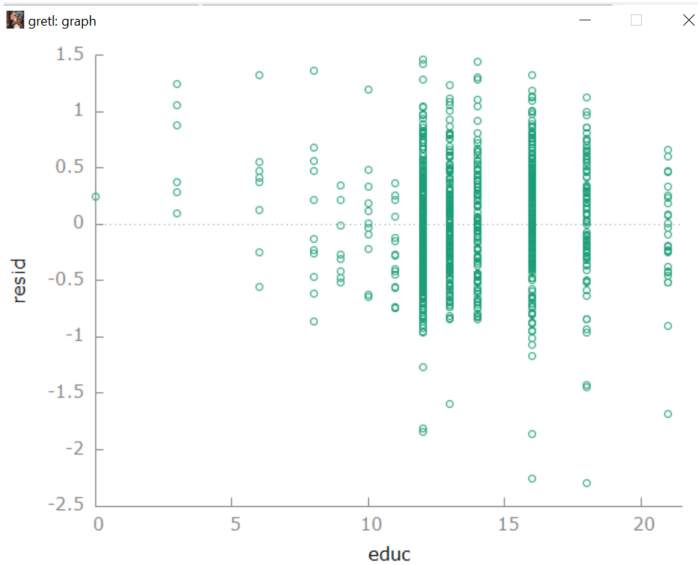


The residual plot suggests the variance of wages for married workers is greater than that for unmarried workers. Thus, there is the evidence of heteroskedasticity.
It probably makes better sense to plot squared residuals against the married variable because in reality, variance is a squared term. However, above figure still shows the change in the dispersion of the data-cloud given the explanatory variable. As we can see, the slope of the fitted line is not horizontal, meaning that there is a heteroskedasticity issue



(d) Plot the least squares residuals against EDUC and against EXPER. What do they suggest?





Both residual plots exhibit a pattern in which the absolute magnitudes of the residuals tend to increase as the values of *EDUC* and *EXPER* increase, although for *EXPER* the increase is not very pronounced. Thus, the plots suggest there is heteroskedasticity with the variance dependent on *EDUC* and possibly *EXPER*.

(e) Test for heteroskedasticity using a Breusch-Pagan test where the variance depends on EDUC, EXPER and MARRIED. What do you conclude at a 5% significance level?

modtest --breusch-pagan

```
? modtest --breusch-pagan

Breusch-Pagan test for heteroskedasticity
OLS, using observations 1-1000
Dependent variable: scaled uhat^2

             coefficient    std. error    t-ratio   p-value
  ---------------------------------------------------------
  const       1.44427       0.767360       1.882    0.0601  *
  educ       -0.0482079     0.0498622     -0.9668   0.3339
  exper      -0.0456217     0.0325651     -1.401    0.1615
  exper2      0.000390635   0.000303371    1.288    0.1982
  experedu    0.00262156    0.00167371     1.566    0.1176
  married     0.247908      0.114282       2.169    0.0303  **

  Explained sum of squares = 52.2061

Test statistic: LM = 26.103073,
with p-value = P(Chi-square(5) > 26.103073) = 0.000085
```

The null and alternative hypotheses are

$$H_0: \text{ errors are homoskedastic}$$
$$H_1: \text{errors are heteroskedastic}$$

With H1 implying the error variance depends on one or more of *EXPER*, *EDUC* or *MARRIED*. The value of the test statistic is 26.1, with P value 0.000085, therefore, we reject the null hypothesis and conclude that heteroskedasticity exists.