Introductory Econometrics
Home Assignment 2
Suggested Solution

by Hieu Nguyen

Fall 2024

Solution of the assignment is to be delivered electronically to **254279@muni.cz** by December 11, 2024 23:59:59 the latest. Late submissions will not be accepted, resulting in zero points.

Form teams of two people, please. Only one team member is supposed to submit the solution with both team members' names and email addresses on the first page of the document. Teams are required to work independently, and any form of plagiarism will be treated accordingly. Please understand that the main advantage of teamwork is the synergy from solving the problems together and the possibility to share and discuss your econometric knowledge with your teammate. It is not about a pure division of tasks. So, please, do cooperate and make sure you both understand all solutions completely.

The text itself can be written in any software of your choice (MS Word, LaTeX, Pages etc.), but the .pdf format [5 MB max, .xls(x) can be attached in .zip] of the final document is required.

Please, name the file ECONOMETRICS_Surname1_Surname2_HA02.pdf.

In your report, please, be clear and reasonably concise, but do explain all essential steps (e.g., important matrices) of your solution/reasoning. Keep in mind that not only the correctness of your answers and interpretations is assessed, but also the text-editing quality is an integral part of your output.

Fingers crossed!
Hieu Nguyen

Problem 1: Wage equation for young males

(4 points: 4 · up to 1 point based on the quality and completeness of the analysis)

Dataset `wage4c.gdt` was used in the 1990s to study the interindustry wage differentials estimated for individuals. It contains 935 observations of monthly wages, study and occupational experience, and family and personal characteristics of young males. Find a specific description of the variables in the dataset.

1. Check the dataset, report the main summary descriptive statistics of the original variables used in the model equation below, and briefly discuss whether everything seems all right. Then construct the dependent variable. State and explain your working hypotheses about signs of the slope coefficients first, then estimate this model in Gretl and report the results:

$$\ln(\text{wage}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{tenure} + \beta_4 \text{married} + \epsilon.$$

   Finally, interpret the estimated coefficients $\hat{\beta}_1$ and $\hat{\beta}_4$.

2. State the null hypothesis that another year of general workforce experience has the same impact on wage as another year of tenure with the current employer. Test this hypothesis at the 10% significance level. What do you conclude? *You can do the test manually or via Gretl.*

3. Test the joint hypothesis that $\beta_2 = \beta_4 = 0$ at the 5% significance level. *First, do the test manually. Then, check your results by conducting the same but automated test in Gretl.* **Note: If you only show the test in Gretl, you won't get any point of this question.**

4. As economists, we might be interested in a potential issue of racial discrimination in the labor market. Enrich the model with three additional dummy variables, estimate it again, and report the results in the usual form:

   $$\ln(\text{wage}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{tenure} + \beta_4 \text{married} + \beta_5 \text{darkskin} + \beta_6 \text{south} + \beta_7 \text{urban} + \nu.$$

   Holding other factors fixed, what is the estimated difference in monthly wage between dark skin and non-dark skin individuals? Is it statistically significant? Show clearly your test *either manually or via Grel.*

Solution:

1. The main summary descriptive statistics:

Variable	Mean	Median	S.D.	Min	Max
wage	958	905	404	115	3080
educ	13.5	12.0	2.20	9.00	18.0
exper	11.6	11.0	4.37	1.00	23.0
tenure	7.23	7.00	5.08	0.00	22.0
married	0.893	1.00	0.309	0.00	1.00

   Everything seems fine, according to a brief inspection of the data and the statistics above. The dataset is large, and there are no missing observations. All observations are non-negative. The sample means, maxima, and minima seem realistic both for 1990s monthly wages in USD and for all variables measured in years. It is also interesting to check the negative correlation between educ and exper, which is relatively high, almost -46%. We also observe that nearly 90% of the sampled individual are married.

```
Model 1: OLS, using observations 1-935
Dependent variable: l_wage

Coefficient Std. Error t-ratio p-value
```

```
const 5.33065 0.114378 46.61 0.0000
educ 0.0753568 0.00643491 11.71 0.0000
exper 0.0141191 0.00333826 4.229 0.0000
tenure 0.0127554 0.00255923 4.984 0.0000
married 0.199171 0.0408196 4.879 0.0000

Mean dependent var 6.779004 S.D. dependent var 0.421144
Sum squared resid 136.4675 S.E. of regression 0.383066
R2 0.176201 Adjusted R2 0.172658
F (4, 930) 49.72908 P-value(F ) 5.96e{38
Log-likelihood 427.0223 Akaike criterion 864.0447
Schwarz criterion 888.2474 Hannan{Quinn 873.2734
```

- all coefficients appear strongly statistically significant at all reasonable levels;
- $\hat{\beta}_1 \approx 0.075$ . . . 1 unit (1 year) change of years of education is associated with circa 7.5% change in the same direction of monthly earnings (USD), ceteris paribus (a log-linear functional form);
- $\hat{\beta}_4 \approx 0.20$ . . . married men earn circa 20% higher salary compared to non-married, ceteris paribus (a log-linear functional form with a dummy).

2. This is an F-test of joint hypotheses/linear restrictions that relates two coefficients.

   The unrestricted model is Model 1.

   We test: $H_0 : \beta_2 = \beta_3$ vs $H_A : \beta_2 \neq \beta_3$ $\Rightarrow$ $F = \frac{(RSS_R - RSS_U)/J}{RSS_U/(n-k-1)} \sim F_{J,n-k-1}$.

   The restricted model is:

$$\ln(\text{wage}) = \beta_0 + \beta_1 \text{educ} + \beta_2(\text{exper} + \text{tenure}) + \beta_4 \text{married} + u.$$

   After transforming the data to a new variable exper plus tenure = exper + tenure, we run the regressions of the restricted model and conduct a standard F-test with $J = 1$ and $k = 4$:

```
Model 3: OLS, using observations 1{935
Dependent variable: l wage

Coefficient Std. Error t-ratio p-value

const 5.34544 0.102405 52.20 0.0000
educ 0.0746565 0.00596462 12.52 0.0000
exper plus tenure 0.0132947 0.00176337 7.539 0.0000
married 0.199477 0.0407860 4.891 0.0000

Mean dependent var 6.779004 S.D. dependent var 0.421144
Sum squared resid 136.4799 S.E. of regression 0.382877
R2 0.176126 Adjusted R2 0.173471
F (3, 931) 66.34246 P-value(F ) 7.00e{39
Log-likelihood 427.0649 Akaike criterion 862.1297
Schwarz criterion 881.4919 Hannan{Quinn 869.5127
```

   We compute from the regression outputs the F-statistic:

$$F \approx \frac{(136.48 - 136.4675)/1}{136.4675/930} \approx 0.085.$$

   The critical value for an F-test is $F_{1,930,0.90} \approx 2.71$.

   Since $F \approx 0.13 < F_{1,930,0.90} \approx 2.71$, we do not reject the H_0 at the given significance level. We can thus conclude that β_2 and β_3 are statistically similar (at the 90% confidence level).

3. Here we test the joint significance of two coefficients because we test for this (incomplete) set of two joint hypotheses using an F-test:

   The unrestricted model is Model 1.

   We test:

$$H_0 : \begin{cases} \beta_2 = 0 \\ \beta_4 = 0 \end{cases} \quad \text{vs} \quad H_A : \begin{cases} \beta_2 \neq 0 \\ \beta_4 \neq 0 \end{cases} \quad \Rightarrow \quad F = \frac{(RSS_R - RSS_U)/J}{RSS_U/(n-k-1)} \sim F_{J,n-k-1}.$$

   The restricted model is:

$$\ln(\text{wage}) = \beta_0 + \beta_1 \text{educ} + \beta_3 \text{tenure} + v.$$

   Estimated output from Gretl:

```
Model 3: OLS, using observations 1{935
Dependent variable: l wage

Coefficient Std. Error t-ratio p-value

const 5.83613 0.0823983 70.83 0.0000
educ 0.0612079 0.00584008 10.48 0.0000
tenure 0.0163803 0.00252771 6.480 0.0000

Mean dependent var 6.779004 S.D. dependent var 0.421144
Sum squared resid 143.0720 S.E. of regression 0.391804
R2 0.136332 Adjusted R2 0.134479
F (2, 932) 73.55931 P-value(F ) 2.18e{30
Log-likelihood 449.1172 Akaike criterion 904.2344
Schwarz criterion 918.7561 Hannan{Quinn 909.7716
```

   We compute from the regression outputs the F-statistic:

$$F \approx \frac{(143.07 - 136.4675)/2}{136.4675/930} \approx 22.5.$$

   The critical value for an F-test is $F_{2,930,0.95} = 3.01$.

   Since $F \approx 22.5$ is much greater than 3.01, we do reject the H_0 of joint insignificance of β_2 and β_4 at the given significance level. We can thus conclude that at least one of the two coefficients is statistically significantly different from 0 (at the 95% confidence level).

   The Gretl output of the test of linear restrictions suggests the same conclusion:

```
Restriction set

1: b[exper] = 0

2: b[married] = 0

Test statistic: F(2, 930) = 22.5043, with p-value = 2.85515e-10

Restricted estimates:

coefficient std. error t-ratio p-value

-------------------------------------------------------
```

```
const 5.83613 0.0823983 70.83 0.0000 ***

educ 0.0612079 0.00584008 10.48 2.26e-24 ***

exper 0.00000 0.00000 NA NA

tenure 0.0163803 0.00252771 6.480 1.48e-10 ***

married 0.00000 0.00000 NA NA

Standard error of the regression = 0.391804
```

4. The Gretl output:

```
Model 4: OLS, using observations 1{935
Dependent variable: l wage

Coefficient Std. Error t-ratio p-value

const 5.39550 0.113225 47.65 0.0000
educ 0.0654307 0.00625040 10.47 0.0000
exper 0.0140430 0.00318519 4.409 0.0000
tenure 0.0117473 0.00245297 4.789 0.0000
married 0.199417 0.0390502 5.107 0.0000
darkskin 0.188350 0.0376666 5.000 0.0000
south 0.0909037 0.0262485 3.463 0.0006
urban 0.183912 0.0269583 6.822 0.0000

Mean dependent var 6.779004 S.D. dependent var 0.421144
Sum squared resid 123.8185 S.E. of regression 0.365471
R2 0.252558 Adjusted R2 0.246914
F (7, 927) 44.74706 P-value(F ) 1.16e{54
Log-likelihood 381.5490 Akaike criterion 779.0979
Schwarz criterion 817.8223 Hannan{Quinn 793.8638
```

We observe a statistically significant negative impact of darkskin. Having dark skin significantly decreasing the monthly salary of young men on average by almost 19% compared to non-dark skin individuals, ceteris paribus, can be interpreted as empirically aggregated evidence for racial discrimination. The conclusion is based on a standard two-sided t-test or simple inspection of the p-value from the Gretl output above.

Problem 2: Wage Equation and Return to Education in 70s

(6 points)

One of the most recent Nobel laureates in Economic Sciences from 2021, David Card, used in his working paper from 1993 wage and education data for a sample of men in the U.S. in 1976 to estimate the return to education. Dataset `card.csv` contains 3,010 observations of hourly wages, schooling and occupational experience, family and personal characteristics, and potential proxies for unobserved personal qualities. Please find a specific description of the variables and the units of measurement in the attached `.txt` file. Be aware that there are missing values in the dataset.

The task of this creative empirical exercise is to develop your own explanatory/predictive model for the determination of individual wage ('wage equation'). The goal is **not to develop the best**

possible model based on the given dataset but to successfully create a **relatively simple but useful and intuitive empirical model that includes the essential variables** while following general suggestions in parts 1. to 6. You should also carefully report the progress of your analysis step by step. If you would like to extend your analysis even further (either in an individual part or in general, e.g., with the multicollinearity analysis or by considering more variables), you are more than welcome to do so.

1. (1.5 pts) Suggest a few (two or three) intuitively the most important explanatory variables for the determination of wage and report and briefly describe their main summary statistics (also include the dependent wage). Discuss suitable functional forms of the variables and estimate the resulting model with OLS. Comment on and interpret the important findings from the result of OLS.

2. (1.5 pt) Suggest two additional potentially important explanatory variables and, taking advantage of the four important variable selection criteria, analyze if they belong to the model. Also, compare your new model to two alternative models with different functional forms between variables.

3. (2 pts) Add from the dataset two or three potentially important intercept dummies and explain why your selection makes sense. Re-estimate the extended model, interpret the newly estimated coefficients, and decide whether the new dummies should remain in the model. Next, add a slope dummy interacting with one of the included quantitative variables, explain your motivation, and interpret the newly added estimated coefficient. Finally, add another interaction term (between two quantitative variables or two dummies), re-estimate, and interpret. What is your resulting model after this step of the analysis? Interpret the overall significance of the regression.

4. (1 pt) Apply the White test for heteroskedasticity. Should we re-calculate the model to obtain heteroskedasticity-robust standard errors? If yes, please do so and interpret your results.

Attached:

1. Dataset: `card.csv`, `wage4c.gdt`

2. Description of variables: `card_description2.txt`

Solution: Students are free to choose regressors and hence, following answers are suggested, not the fixed results. Grades to students will be delivered case by case.

1. We choose the two most important regressors: *educ* and *exper*, and we use the quadratic form for *exper* as we experienced before that we might expect concave returns to cumulative experience. Also note that we use the *log(wage)* as the dependent variable as we can expect an increasing rate of change (wage usually grows in percentage terms while *exper* in years, i.e., unit terms). As we expected, all of the coefficients are statistically significant, but the adjusted R^2 is rather low, while the return to *experience* is really concave.

```
log(wage) ~ educ + exper + exper^2

Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.4685404  0.0686899  65.054 < 2e-16 ***
educ         0.0931707  0.0035802  26.024 < 2e-16 ***
exper        0.0897828  0.0070636  12.711 < 2e-16 ***
I(exper^2)  -0.0024859  0.0003377  -7.361 2.35e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3982 on 3006 degrees of freedom
Multiple R-squared:  0.1958,   Adjusted R-squared:  0.195
F-statistic:  244 on 3 and 3006 DF,  p-value: < 2.2e-16
```

2. Adding two additional variables: *IQ* and *Nearc4*. A higher *IQ* should help us get better jobs, and we could have leveraged our work-related skills and knowledge by having a college nearby.

```
    log(wage) ~ educ + exper + exper^2 + IQ + nearc4

                Estimate Std. Error t value Pr(>|t|)
    (Intercept)  3.9840038 0.0789789  50.444 < 2e-16 ***
    educ         0.0757135 0.0037452  20.216 < 2e-16 ***
    exper        0.0967456 0.0069212  13.978 < 2e-16 ***
    I(exper^2)  -0.0028335 0.0003307  -8.569 < 2e-16 ***
    IQ           0.0062390 0.0005607  11.128 < 2e-16 ***
    nearc4       0.0888423 0.0153913   5.772 8.62e-09 ***

    Residual standard error: 0.3877 on 3004 degrees of freedom
    Multiple R-squared: 0.2383, Adjusted R-squared: 0.237
    F-statistic: 187.9 on 5 and 3004 DF, p-value: < 2.2e-16
```

All variables are strongly significant. 0.6 pct wage return to *IQ* looks realistic; same for the *nearc4* dummy (9 pct). The coefficients on *educ* went down from 0.093 to 0.076. For *experience* the change is not that high. The adjusted R^2 increases from 0.195 to 0.237, which is a considerable relative change but still includes new explanatory power. Considering the four important selection criteria, the *theory/intuition* suggests both variables should be included, as discussed above. We also most likely observe an *omitted variable bias reduction* as both variables are positively correlated with *educ*, which coefficient went down. The adjusted R^2 increases, and both new variables are strongly statistically significant. Both variables thus most likely belong to the model.

Regarding different functional forms, we will further check whether there is a diminishing return to *IQ* and if the relationship between *wage* and *educ* cannot be better described via an elasticity (log-log).

```
lm(formula = I(log(wage)) ~ educ + exper + I(exper^2) + IQ + I(IQ^2) + nearc4,
    data = wage_data_filtered)

                Estimate Std. Error t value Pr(>|t|)
    (Intercept)  2.786e+00 2.473e-01  11.265 < 2e-16 ***
    educ         7.773e-02 3.750e-03  20.725 < 2e-16 ***
    exper        9.567e-02 6.896e-03  13.874 < 2e-16 ***
    I(exper^2)  -2.778e-03 3.295e-04  -8.430 < 2e-16 ***
    IQ           3.060e-03 4.799e-04   6.376 2.09e-10 ***
    I(IQ^2)     -1.233e-05 2.413e-05  -5.111 3.41e-07 ***
    nearc4       8.687e-02 1.533e-02   5.666 1.60e-08 ***

    Residual standard error: 0.386 on 3003 degrees of freedom
    Multiple R-squared: 0.2448, Adjusted R-squared: 0.2433
    F-statistic: 162.3 on 6 and 3003 DF, p-value: < 2.2e-16

log(wage) ~ log(educ) + exper + exper^2 + IQ + nearc4

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
    (Intercept)  2.8523436 0.1219714  23.385 < 2e-16 ***
    I(log(educ)) 0.8388992 0.0435142  19.279 < 2e-16 ***
    exper        0.0777618 0.0068672  11.324 < 2e-16 ***
    I(exper^2)  -0.0019325 0.0003374  -5.728 1.12e-08 ***
    IQ           0.0069262 0.0005550  12.481 < 2e-16 ***
    nearc4       0.0887111 0.0154835   5.729 1.11e-08 ***

    Residual standard error: 0.3898 on 3004 degrees of freedom
    Multiple R-squared: 0.2299, Adjusted R-squared: 0.2286
    F-statistic: 179.4 on 5 and 3004 DF, p-value: < 2.2e-16
```

In both models, all coefficients are strongly significant, while only for the first model does the adjusted R^2 increase. *IQ* thus seems to have rather diminishing returns similarly to *exper*, which

is intuitive, as higher-*IQ* people will, on average, tend to have higher income, but not necessarily and steadily linearly increasing with *IQ*. It can also be observed visually in Figure 1. On the other hand, a lower R^2 of the second model supports our original semi-log functional relationship between *wage* and *educ*. Recall that two models can only be compared using R^2 if they have the same dependent variable.

3. Add two intercept dummies: take the best model so far and add *south* and *black* as dummy variables. Black people were less educated in the 70s'. Also, the South was a less important region in the past. We do expect negative signs for both.

```
log(wage) ~ educ + exper + exper^2 + IQ + IQ^2
    + nearc4 + black + south

              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.689e+00 2.608e-01  14.147 < 2e-16 ***
educ          7.397e-02 3.686e-03  20.070 < 2e-16 ***
exper         8.333e-02 6.781e-03  13.025 < 2e-16 ***
I(exper^2)  -2.483e-03 3.235e-04  -7.676 2.20e-14 ***
IQ            1.917e-02 4.890e-03   3.921 9.02e-05 ***
I(IQ^2)     -8.172e-05 2.417e-05  -3.382 0.00073 ***
nearc4        6.301e-02 1.529e-02   4.121 3.87e-05 ***
black       -1.217e-01 2.094e-02  -5.810 6.92e-09 ***
south       -1.313e-01 1.541e-02  -8.518 < 2e-16 ***

Residual standard error: 0.3779 on 3001 degrees of freedom
Multiple R-squared: 0.277,   Adjusted R-squared: 0.2751
F-statistic: 143.7 on 8 and 3001 DF,  p-value: < 2.2e-16
```

We expected these signs. The new coefficients are strongly significant, and the adjusted R^2 improved. Being *black* and coming from the South was not economically beneficial at that time. Especially the estimated coefficients on *IQ* and *nearc4* have changed markedly, which suggests another reduction of the omitted variable bias. The new variables will thus remain in the model.

Let us additionally insert (*black* * *educ*) representing a slope dummy, i.e., an interaction term between a dummy and a quantitative variable. The motivation is that a significant parameter would indicate discrimination in the return to education.

```
log(wage) ~ educ + exper + exper^2 + IQ + IQ^2
    + nearc4 + black + south + black * educ

              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.881e+00 2.675e-01  14.512 < 2e-16 ***
educ          6.910e-02 3.993e-03  17.303 < 2e-16 ***
exper         8.552e-02 6.830e-03  12.521 < 2e-16 ***
I(exper^2)  -2.321e-03 3.271e-04  -7.094 1.62e-12 ***
IQ            1.669e-02 4.947e-03   3.374 0.000759 ***
I(IQ^2)     -6.867e-05 2.449e-05  -2.804 0.005073 **
nearc4        6.138e-02 1.528e-02   4.018 6.01e-05 ***
black       -8.247e-02 1.824e-02  -4.515 6.57e-06 ***
south       -1.306e-01 1.539e-02  -8.488 < 2e-16 ***
I(black * educ)  2.024e-02 6.441e-03   3.143 0.001699 **

Residual standard error: 0.3773 on 3000 degrees of freedom
Multiple R-squared: 0.2794,   Adjusted R-squared: 0.2772
F-statistic: 129.2 on 9 and 3000 DF,  p-value: < 2.2e-16
```

Adjusted R^2 has slightly increased, and the new coefficient is statistically significant. The new interaction term has a positive effect, which is counterintuitive, but conversely, the estimate on *educ* has decreased from 0.074 to 0.069, similar to *black* from -0.12 to -0.37. Thus, the model

now captures the returns to education in greater detail and isolates the impact of *educ* alone, *black* alone, which is much stronger than previously, and the interaction impact of the two, which balances their drop.

We now add (*nearc4* * *black*) as an interaction term between two dummies for a potentially similar reason as before, but now w.r.t. the impact of the proximity to a 4-year college.

```
log(wage) ~ educ + exper + exper^2 + IQ + IQ^2
    + nearc4 + black + south + educ * black + nearc4 * black

                Estimate Std. Error t value Pr(>|t|)
    (Intercept)  3.868e+00 2.698e-01  14.337 < 2e-16 ***
    educ         6.903e-02 3.999e-03  17.261 < 2e-16 ***
    exper        8.558e-02 6.834e-03  12.524 < 2e-16 ***
    I(exper^2)  -2.325e-03 3.274e-04  -7.101 1.54e-12 ***
    IQ           1.690e-02 4.981e-03   3.394 0.000699 ***
    I(IQ^2)     -6.966e-05 2.463e-05  -2.828 0.004718 **
    nearc4       6.463e-02 1.759e-02   3.674 0.000243 ***
    black       -3.681e-01 8.328e-02  -4.420 1.02e-05 ***
    south       -1.305e-01 1.540e-02  -8.472 < 2e-16 ***
    I(black * educ)  2.060e-02 6.514e-03   3.163 0.001579 **
    I(nearc4 * black) -1.295e-02 3.479e-02  -0.372 0.709736

    Residual standard error: 0.3774 on 2999 degrees of freedom
    Multiple R-squared: 0.2794,   Adjusted R-squared: 0.277
    F-statistic: 116.3 on 10 and 2999 DF,  p-value: < 2.2e-16
```

In this setup, the coefficient of (*nearc4* * *black*) is insignificant, estimated negative, and the adjusted R^2 decreased. The only difference worth mentioning is related to a slight increase in the effect of *nearc4*. Considering the four important selection criteria, there is no strong intuition behind this new term; the change in *nearc4* is due to multicollinearity, not an *OVB* reduction; and also, the *technical criteria* (adjusted R^2, t test) suggest this new dummy interaction term is not important for explaining wage differentials.

The resulting model is thus the first one with (*black* * *educ*) only. The p-value for the null hypothesis of the overall insignificance of the regression is practically zero ($< 2.2e-16$), clearly a rejection of the null (which is also obvious from individual t stats).

4. White test for heteroskedasticity: Let us use the model from the previous part, append residuals to the dataset, and specify the White test as follows. *It is also accepted if you use Gretl directly for heteroskedasticity test.*

```
res^2 ~ educ + exper + exper^2 + IQ + IQ^2
    + nearc4 + black + south + (black*educ)^2
    + educ^2 + exper^2 + exper^4 + I(IQ^2) + I(IQ^4)
    + nearc4^2) + black^2 + south^2 + (educ*black)^2
```

We do not report the estimated model here because of its size, but some of its parameters appear strongly significant, e.g., on *educ*. The resulting p-value for the F test of the null hypothesis of joint insignificance is negligible (1.556e - 05), and similarly for the LM test (0.000175). We thus reject the null of homoskedasticity at any standard level. For a check of correct coding, the calculated p-value for the F test must be the same as the p-value for testing the overall insignificance of the regression that is directly reported with the estimated model.

We should recalculate the robust standard errors:

```
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     3.8815e+00 3.1682e-01  12.2513 < 2.2e-16 ***
educ            6.9100e-02 4.1287e-03  16.7366 < 2.2e-16 ***
exper           8.5516e-02 6.8208e-03  12.5376 < 2.2e-16 ***
```

```
I(exper^2)     -2.3207e-03 3.2320e-04  -7.1805 8.718e-13 ***
IQ              1.6689e-02 5.9075e-03   2.8250  0.004759 **
I(IQ^2)        -6.8671e-05 2.8729e-05  -2.3903  0.016896 *
nearc4          6.1381e-02 1.5065e-02   4.0743 4.735e-05 ***
black          -3.7235e-01 8.0756e-02  -4.6108 4.179e-06 ***
south          -1.3063e-01 1.5660e-02  -8.3419 < 2.2e-16 ***
I(black * educ) 2.0241e-02 6.1768e-03   3.2770  0.001061 **
```

and compare them with the original model:

```
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     3.881495e+00 2.674754e-01  14.511600 3.615748e-46
educ            6.909963e-02 3.993452e-03  17.303232 5.153031e-64
exper           8.551639e-02 6.830091e-03  12.520533 4.279323e-35
I(exper^2)     -2.320718e-03 3.271440e-04  -7.093871 1.620077e-12
IQ              1.668900e-02 4.946838e-03   3.373638 7.512332e-04
I(IQ^2)        -6.867056e-05 2.448669e-05  -2.804404 5.073420e-03
nearc4          6.138072e-02 1.527587e-02   4.018148 6.009867e-05
black          -3.723512e-01 8.246550e-02  -4.515236 6.568315e-06
south          -1.306338e-01 1.539129e-02  -8.487513 3.269500e-17
I(black * educ) 2.024143e-02 6.441038e-03   3.142559 1.691116e-03
```

We observe that for some variables, the robust standard errors are larger, and thus, the t statistics have decreased, especially in the cases of *IQ* and *educ*. This, however, does not threaten the specification of our model as we have a large sample size. Nevertheless, we indeed had a biased estimation of the error variance before.