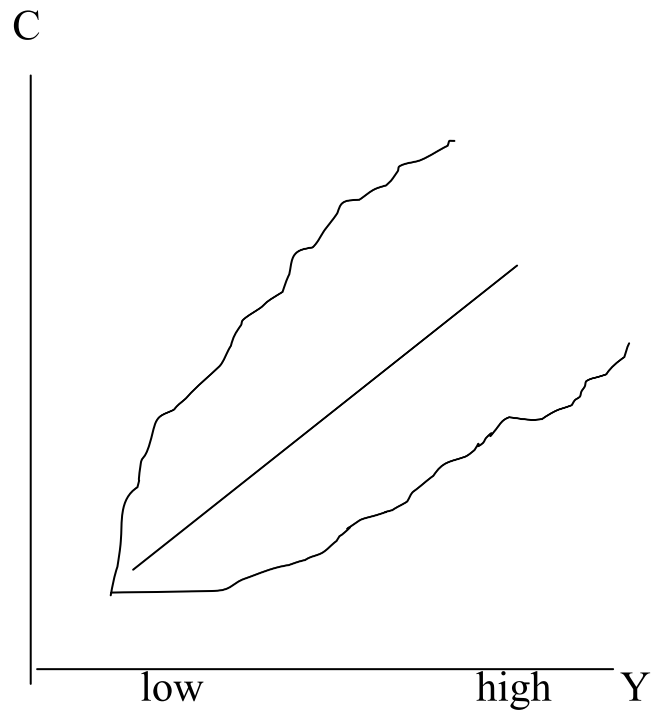
8. HETEROSKEDASTICITY

[1] Definition of Heteroskedasticity (HET)

(1) Model: $y_t = x_{t\bullet}'\beta + \varepsilon_t$.

We say that HET exists if $var(\varepsilon_t) \neq var(\varepsilon_s)$, $t \neq s$.

(2) Example



In general,

$$Cov(\varepsilon) = \sigma^2\Omega = \sigma^2 \begin{pmatrix} \omega_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \omega_2^2 & 0 & \cdots & 0 \\ 0 & 0 & \omega_3^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \omega_T^2 \end{pmatrix}.$$

→ It happens if the ε_t are stochastically independent but not identically distributed.

[2] Properties of OLS under HET

Assumptions:

- All of SIC except SIC.6 (the homoskedasticity assumption).
- Or all of WIC except WIC.5 (the homoskedasticity assumption).

Theorem:

$\hat{\beta}$ is unbiased (consistent) but inefficient. Further, the usual t or F tests using $s^2(X'X)^{-1}$ as estimated $Cov(\hat{\beta})$ are all invalid.

Theorem:

$$Cov(\hat{\beta}) = (X'X)^{-1}X'(\sigma^2\Omega)X(X'X)^{-1}.$$

Theorem: White (1980, Econometrica)

Let $\hat{\Delta} = X'diag(e_1^2, e_2^2, ..., e_T^2)X = \Sigma_t e_t^2 x_{t\bullet} x_{t\bullet}'$. Then,

$$p\lim \frac{\hat{\Delta}}{T} = \lim \frac{1}{T} X'(\sigma^2\Omega)X .$$

Comment:

- Here, the e_t are the OLS residuals.
- We can estimate $X'(\sigma^2\Omega)X$ by $\hat{\Delta}$, if sample is large.

Definition: (White-corrected covariance matrix of OLS estimator)

Estimated $Cov(\hat{\beta}) = (X'X)^{-1}\hat{\Delta}(X'X)^{-1}$.

[3] TESTING HYPOTHESES BASED ON OLS

- Testing linear restrictions:
  - Under Ho: $R\beta = r$ (m restrictions),

    $$W_T = \left(R\hat{\beta} - r\right)' \left(R\left(Estimated\ Cov(\hat{\beta})\right)R'\right)^{-1} \left(R\hat{\beta} - r\right) \to_d \chi^2(m).$$

    Or,

    $W_T / m \approx F(m, T-k)$.

    (CAUTION!) The F-test of the form

    $$[(SSE_R - SSE_{UR})/m]/[SSE_{UR}/(T-k)]$$

    should not ne used.
  - Under H_o: $\beta_j = \beta_j^*$,

$$t_j = \frac{\hat{\beta}_j - \beta_j^*}{se(\hat{\beta}_j)} \rightarrow_d N(0,1).$$

where $se(\hat{\beta}_j) = \sqrt{j'th\ diagonal\ of\ (X'X)^{-1}\hat{\Delta}(X'X)^{-1}}$.

- Under H_o: Rβ = r (1 restrictions),
- $t = \frac{R\hat{\beta} - r}{\sqrt{RCov(\hat{\beta})R'}} \rightarrow_d N(0,1)$.

- Testing nonlinear restrictions:
  - H_o: w(β) = $0_{m\times 1}$ (multiple restrictions):
    - Define $W(\beta) = \frac{\partial w(\beta)}{\partial \beta'}$.
    - $W_T = w(\hat{\beta})'\left[W(\hat{\beta})Cov(\hat{\beta})W(\hat{\beta})'\right]^{-1} w(\hat{\beta}) \rightarrow_d \chi^2(m)$.
  - H_o: w(β) = 0 (single restriction):
    - Use the Wald test or t = $\frac{w(\hat{\beta})}{\sqrt{W(\hat{\beta})Cov(\hat{\beta})W(\hat{\beta})'}}$.

[4] TESTING HET

(1) Breusch-Pagan Test (BP)

- Assume that $\sigma^2\omega_t^2 = h(z_{t\bullet}'\gamma)$, where the functional form of h need not be known, $z_{t\bullet} = (z_{t1}, \ldots, z_{tp})'$, $\gamma = [\gamma_1, \ldots, \gamma_p]'$ and $z_{t1} = 1$.

Comment:

- Usually, $z_{t\bullet} = x_{t\bullet}$.
- $h(z_{t\bullet}'\gamma)$ would be like $(z_{t\bullet}'\gamma)^2$, or $\exp(z_{t\bullet}'\gamma)$
- Observe that if $\gamma_2 = \ldots = \gamma_p = 0$ then $\sigma^2\omega_t^2 = h(\gamma_1)$ is constant so that there is no HET.
- H_o: no HET $\leftrightarrow$ Ho: $\gamma_2 = \ldots = \gamma_p = 0$.

LM Test Procedure:

STEP 1: Do OLS on $y_t = x_{t\bullet}'\beta + \varepsilon_t$ and get residuals, e_t.

STEP 2: Compute $\hat{\sigma}_{ML}^2 = \frac{1}{T}\Sigma_t e_t^2$ = SSE/T or s^2.

STEP 3: Do OLS on $\frac{e_t^2}{\hat{\sigma}_{ML}^2} = z_{t\bullet}'\gamma + error$ and get SSR.

STEP 4: BP = SSR/2 $\rightarrow \chi^2(df=p-1)$.

(2) Koenker Test (1981, Journal of Econometrics)

STEP 1: Do OLS on $y_t = x_{t\bullet}'\beta + \varepsilon_t$ and get residuals, e_t.

STEP 2: Do OLS on $e_t^2 = z_{t\bullet}'\gamma$ + error and get R^2.

STEP 3: Koenker = $T \times R^2 \to \chi^2(df=p-1)$.

Comment:

If the ε_t are normal, then Koenker and BP are asymptotically identical.

If the ε_t are not normal, Koenker test is O.K., but not BP.

(3) White's Information Matrix Test (1980, Econometrica)

$z_{t\bullet}$ = vector of cross products of variables in $x_{t\bullet}$.

Example:

$x_{t\bullet}' = [1, x_{t2},, x_{t3}] \to z_{t\bullet}' = [1, x_{t2}, x_{t3}, x_{t2}^2, x_{t3}^2, x_{t2}x_{t3}]$.

If x_{t3} is a dummy variable, $z_{t\bullet}' = [1, x_{t2}, x_{t3}, x_{t2}^2, x_{t2}x_{t3}]$ (since $x_{t3}^2 = x_{t3}$).

STEP 1: Do OLS on $y_t = x_{t\bullet}'\beta + \varepsilon_t$ and get residuals, e_t.

STEP 2: Do OLS on $e_t^2 = z_{t\bullet}'\gamma$ + error and get R^2.

STEP 3: White = $T \times R^2 \to \chi^2(df=p-1)$.

$\to$ Similar to Koenker.

[5] GLS

(1) $y = X\beta + \varepsilon$, with $Cov(\varepsilon) = \sigma^2\Omega$.

- For GLS, transform the model into $Vy = VX\beta + V\varepsilon$, where $V'V = \Omega^{-1}$.

(2) Cases where Ω is known: (We can use GLS)

- Suppose $\sigma^2\Omega = \sigma^2 diag(\omega_1^2, \ldots , \omega_T^2)$.

  In this case, $\Omega = diag(\omega_1^2, \ldots. , \omega_T^2)$.

- $\Omega^{-1} = diag(1/\omega_1^2, \ldots , 1/\omega_T^2) \rightarrow V = diag(1/\omega_1, \ldots , 1/\omega_T)$.

$$V\varepsilon = \begin{pmatrix} \frac{1}{\omega_1} & 0 & \ldots & 0 \\ 0 & \frac{1}{\omega_2} & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \frac{1}{\omega_T} \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix} = \begin{pmatrix} \frac{\varepsilon_1}{\omega_1} \\ \frac{\varepsilon_2}{\omega_2} \\ \vdots \\ \frac{\varepsilon_T}{\omega_T} \end{pmatrix}.$$

- $E(V\varepsilon) = 0$, $Cov(V\varepsilon) = \sigma^2 I_T$.

(3) Cases where $\Omega = diag(\omega_1^2, \ldots , \omega_T^2)$ is unknown:

- Have to use FGLS, that is, we need to replace ω_t^2 by $\hat{\omega}_t^2$.
- To do GLS, you can also use the estimated $\sigma^2\omega_t^2$ (say, $\widehat{\sigma^2\omega_t^2}$).

Example 1:

- Consider:

$$y_t = x_{t1}\beta_1 + x_{t2}\beta_2 + ... + x_{tk}\beta_k + \varepsilon_t \text{ with } var(\varepsilon_t) = \sigma^2\omega_t^2.$$

- Assume we know ω_t^2's. Then, the transformed model is given by:

$$y_t^* = x_{t1}^*\beta_1 + x_{t2}^*\beta_2 + ... + x_{tk}^*\beta_k + \varepsilon_t^*, \qquad (*)$$

  where $y_t^* = y_t/\omega_t$; $x_{tj}^* = x_{tj}/\omega_t$; and $\varepsilon_t^* = \varepsilon_t/\omega_t$. GLS = OLS on (*).

Example 2:

- $y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$, where $var(\varepsilon_t) = \sigma^2 x_t^2$ ($\omega_t = x_t$)
- The transformed equation is given by:

$$(y_t/x_t) = \beta_1(1/x_t) + \beta_2 + (\varepsilon_t/x_t).$$

Example 3:

- $y_t = x_{t\bullet}'\beta + \varepsilon_t$, where β is a $k\times 1$ vector and $var(\varepsilon_t) = \sigma^2(x_{t\bullet}'\beta)^2$.

  $\rightarrow \omega_t = x_{t\bullet}'\beta$.

- Do FGLS ($\hat{\omega}_t = x_{t\bullet}'\hat{\beta} = \hat{y}_t$).

  STEP 1: Do OLS and get $\hat{\beta}$ and $\hat{y}_t$.

  STEP 2: The transformed equation is given:

$$y_t / \hat{y}_t = (x_{t\bullet} / \hat{y}_t)'\beta + \varepsilon_t / \hat{y}_t \qquad (*)$$

  STEP 3: FGLS = OLS on (*).

Comments on Example 3:

- The resulting estimator is a feasible GLS.
- More efficient estimator can be obtained by MLE.

  (Amemiya, 1973, JASA)

Example 4: Grouped Data

- $y_{ij} = \beta_1 + \beta_2 x_{ij} + \varepsilon_{ij}$,

  where "i" denotes "states," "j" denotes "persons," and $\text{var}(\varepsilon_{ij}) = \sigma^2$.

- Suppose we have data on $\overline{y}_i = \frac{1}{n_i}\Sigma_{j=1}^{n_i} y_{ij}$ and $\overline{x}_i = \frac{1}{n_i}\Sigma_{j=1}^{n_i} x_{ij}$.

- Consider the model:

$$\overline{y}_i = \beta_1 + \beta_2 \overline{x}_i + \overline{\varepsilon}_i .$$

- Note that $\text{var}(\overline{\varepsilon}_i) = \frac{\sigma^2}{n_i} = \sigma^2 \omega_i^2$ ($\omega_i = \frac{1}{\sqrt{n_i}}$). Transform the model into:

$$\sqrt{n_i}\,\overline{y}_i = \beta_1\left(\sqrt{n_i}\right) + \beta_2\left(\sqrt{n_i}\,\overline{x}_i\right) + \left(\sqrt{n_i}\,\varepsilon_i\right).$$

Example 6: (Random Coefficient Model: See Judge, et al (p. 434))

- $y_t = \alpha + \beta_t x_t + \varepsilon_t$; $\beta_t = \beta + v_t$,

  ε_t iid $N(0,\sigma^2)$, v_t iid $N(0,\sigma_v^2)$ and $cov(\varepsilon_t,v_t) = 0$.

- This model can be written as:

$$y_t = \alpha + \beta x_t + (\varepsilon_t + v_t x_t) \,, \qquad (*)$$

  where $var(\varepsilon_t+v_t x_t) = \sigma^2 + \sigma_v^2 x_t^2$.

STEP 1: Do OLS on (*) get e_t

STEP 2: Do OLS on $e_t^2 = \sigma^2 + \sigma_v^2 x_t^2$ + error, and get

$\hat{e}_t^2 = \hat{\sigma}^2 + \hat{\sigma}_v^2 x_t^2$ (fitted value).

STEP3: Define $\hat{\omega}_t = \sqrt{\hat{e}_t^2}$.

STEP 4: Do OLS on $\frac{y_t}{\hat{\omega}_t} = \left(\frac{1}{\hat{\omega}_t}\right)\alpha + \left(\frac{x_t}{\hat{\omega}_t}\right)\beta + error$.

<EXAMPLE>

- Data: (WAGE2.WF1 or WAGE2.TXT – from Wooldridge's website)

  # of observations (T): 935

1. wage	monthly earnings
2. hours	average weekly hours
3. IQ	IQ score
4. KWW	knowledge of world work score
5. educ	years of education
6. exper	years of work experience
7. tenure	years with current employer
8. age	age in years
9. married	=1 if married
10. black	=1 if black
11. south	=1 if live in south
12. urban	=1 if live in SMSA
13. sibs	number of siblings
14. brthord	birth order
15. meduc	mother's education
16. feduc	father's education
17. lwage	natural log of wage

Example:

- $\log(wage) = \beta_1 + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 iq + \varepsilon_t$.

Dependent Variable: LWAGE
Method: Least Squares
Sample: 1 935
Included observations: 935

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.213965	0.132947	39.21823	0.0000
EDUC	0.057326	0.007388	7.758936	0.0000
EXPER	0.015714	0.013301	1.181414	0.2377
EXPER^2	0.000165	0.000557	0.295494	0.7677
IQ	0.005787	0.000980	5.903960	0.0000

R-squared	0.162323	Mean dependent var	6.779004
Adjusted R-squared	0.158720	S.D. dependent var	0.421144
S.E. of regression	0.386279	Akaike info criterion	0.940818
Sum squared resid	138.7665	Schwarz criterion	0.966704
Log likelihood	-434.8326	F-statistic	45.05323
Durbin-Watson stat	1.812399	Prob(F-statistic)	0.000000

Genr res2 = resid^2.

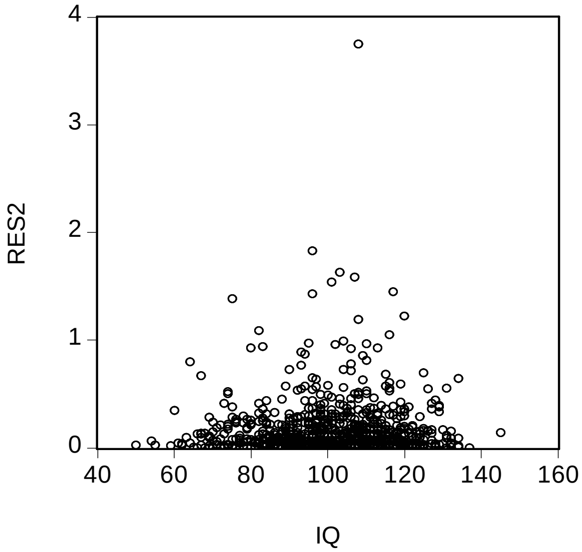
view/residualtests/White Heteroskedasticity (cross terms)

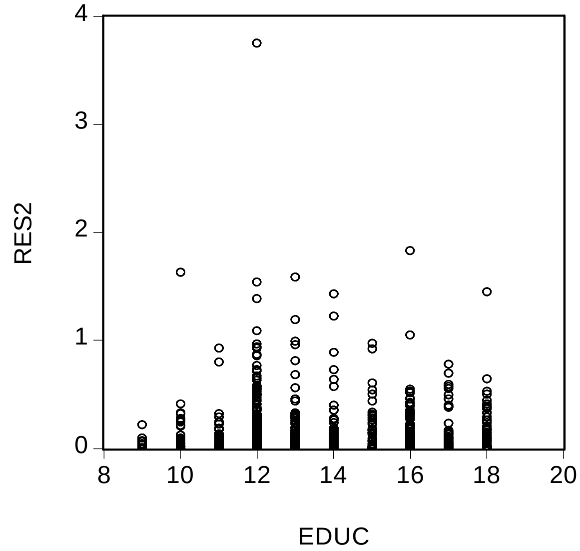
F-statistic	2.194008	Probability	0.008333
Obs*R-squared	28.08589	Probability	0.008803

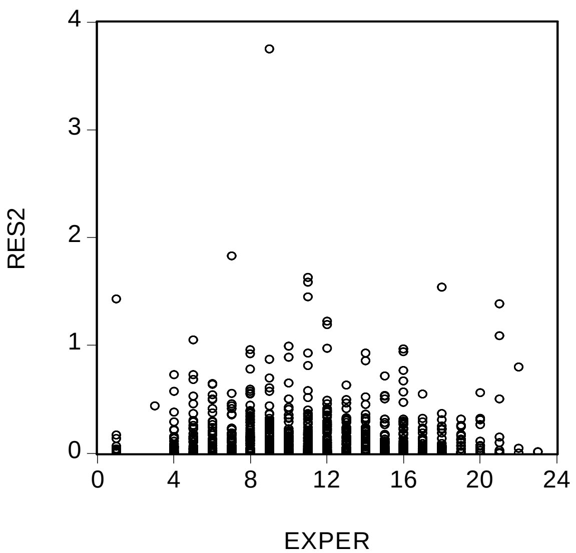
Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Sample: 1 935
Included observations: 935

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.408129	0.629412	-0.648429	0.5169
EDUC	0.046608	0.068175	0.683644	0.4944
EDUC^2	0.004057	0.002469	1.642811	0.1008
EDUC*EXPER	-0.020141	0.006145	-3.277741	0.0011
EDUC*(EXPER^2)	0.000996	0.000282	3.525729	0.0004
EDUC*IQ	-0.000707	0.000386	-1.832457	0.0672
EXPER	0.062635	0.090475	0.692286	0.4889
EXPER^2	-0.001139	0.008320	-0.136919	0.8911
EXPER*(EXPER^2)	-0.000309	0.000458	-0.674600	0.5001
EXPER*IQ	0.002239	0.000762	2.936757	0.0034
(EXPER^2)^2	1.13E-05	9.84E-06	1.148499	0.2511
(EXPER^2)*IQ	-0.000109	3.16E-05	-3.463166	0.0006
IQ	0.001611	0.007678	0.209767	0.8339
IQ^2	-9.17E-06	3.24E-05	-0.282876	0.7773

R-squared	0.030038	Mean dependent var	0.148413
Adjusted R-squared	0.016347	S.D. dependent var	0.251355
S.E. of regression	0.249292	Akaike info criterion	0.074475
Sum squared resid	57.23683	Schwarz criterion	0.146953
Log likelihood	-20.81686	F-statistic	2.194008
Durbin-Watson stat	2.033151	Prob(F-statistic)	0.008333


RES2
IQ
0
1
2
3
4
40
60
80
100
120
140
160


RES2
EDUC
0
1
2
3
4
8
10
12
14
16
18
20


RES2
EXPER
0
1
2
3
4
0
4
8
12
16
20
24

OLS regression results with White correction

Dependent Variable: LWAGE
Method: Least Squares
Sample: 1 935
Included observations: 935
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.213965	0.130678	39.89930	0.0000
EDUC	0.057326	0.007919	7.238736	0.0000
EXPER	0.015714	0.013952	1.126222	0.2604
EXPER^2	0.000165	0.000591	0.278554	0.7806
IQ	0.005787	0.000984	5.882152	0.0000

R-squared	0.162323	Mean dependent var	6.779004
Adjusted R-squared	0.158720	S.D. dependent var	0.421144
S.E. of regression	0.386279	Akaike info criterion	0.940818
Sum squared resid	138.7665	Schwarz criterion	0.966704
Log likelihood	-434.8326	F-statistic	45.05323
Durbin-Watson stat	1.812399	Prob(F-statistic)	0.000000

- Assume:

  $\log(wage) = \beta_1 + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 iq + \varepsilon_t,$

  with $var(\varepsilon_t) = \sigma^2 educ \char`^ 2$

GLS under this assumption:

Dependent Variable: LWAGE/EDUC
Method: Least Squares
Sample: 1 935
Included observations: 935
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
1/EDUC	5.233690	0.129050	40.55551	0.0000
C	0.058700	0.008460	6.938815	0.0000
EXPER/EDUC	0.010861	0.013878	0.782646	0.4340
EXPER^2/EDUC	0.000335	0.000589	0.569002	0.5695
IQ/EDUC	0.005709	0.001017	5.613529	0.0000

R-squared	0.853839	Mean dependent var	0.514690
Adjusted R-squared	0.853211	S.D. dependent var	0.077082
S.E. of regression	0.029532	Akaike info criterion	-4.20133
Sum squared resid	0.811109	Schwarz criterion	-4.17544
Log likelihood	1969.120	F-statistic	1358.217
Durbin-Watson stat	1.775698	Prob(F-statistic)	0.000000

view/residualtests/White Heteroskedasticity (cross terms)

White Heteroskedasticity Test:

F-statistic	5.075883	Probability	0.000000
Obs*R-squared	62.51085	Probability	0.000000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Sample: 1 935
Included observations: 935
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.007083	0.005115	1.384791	0.1665
1/EDUC	0.062730	0.082521	0.760168	0.4473
(1/EDUC)^2	-1.057654	0.433457	-2.440043	0.0149
(1/EDUC)*(EXPER/EDUC)	0.113311	0.080599	1.405853	0.1601
(1/EDUC)*(EXPER^2/EDUC)	-0.004369	0.008901	-0.490862	0.6236
(1/EDUC)*(IQ/EDUC)	0.010280	0.006945	1.480104	0.1392
EXPER/EDUC	-0.024588	0.008240	-2.983905	0.0029
(EXPER/EDUC)*(EXPER^2/EDUC)	-0.000237	0.000498	-0.476583	0.6338
(EXPER/EDUC)*(IQ/EDUC)	0.002399	0.000793	3.025024	0.0026
EXPER^2/EDUC	0.001223	0.000433	2.821980	0.0049
(EXPER^2/EDUC)^2	1.01E-05	1.04E-05	0.966306	0.3341
(EXPER^2/EDUC)*(IQ/EDUC)	-0.000119	3.75E-05	-3.182746	0.0015
IQ/EDUC	-0.001486	0.000593	-2.506438	0.0124
(IQ/EDUC)^2	-4.15E-06	2.27E-05	-0.182452	0.8553

R-squared	0.066857	Mean dependent var	0.000867
Adjusted R-squared	0.053685	S.D. dependent var	0.001597
S.E. of regression	0.001553	Akaike info criterion	-10.08214
Sum squared resid	0.002222	Schwarz criterion	-10.00966
Log likelihood	4727.399	F-statistic	5.075883
Durbin-Watson stat	2.001466	Prob(F-statistic)	0.000000