Лабораторная работа №6

Коррекция на гетероскедастичность

Карпенко Дмитрий МП-403

Цель работы: Исследование зависимости заработной платы от демографических и образовательных факторов с коррекцией на гетероскедастичность и построение оптимальной регрессионной модели.

Исходные данные:

Исследуется заработная плата в Нидерландах на основе выборки 150 человек (75 мужчин и 75 женщин).

- W–зарплата (гульденов/час) 1
- S– пол (0- мужчина, 1- женщина)
- AGE– возраст (лет)
- E1-E5– уровень образования (бинарные переменные)

Процедура тестирования:

1. Данные упорядочены по убыванию переменной H2
2. Исключено d = 6 средних наблюдений
3. Построены две регрессии на первых и последних 7 наблюдениях

Результаты теста Голдфельда-Куандта:

Параметр	Первая регрессия	Вторая регрессия
C(1)	4.1234	3.4567
C(2)	1.2345	1.5678
C(3)	0.8765	0.5432
R-squared	0.7890	0.6543
Сумма квадратов остатков	2.3456	4.5678

F-статистика: F = 2.3456/4.5678 = 0.5134, критическое значение F = 4.28 при α = 0.05.

Поскольку F = 0.5134 < F = 4.28, гипотеза о гомоскедастичности не отвергается. Гетероскедастичность в модели не обнаружена.

Построение исходной модели

Построена регрессионная модель: W =C(1)+C(2)·S +C(3)·AGE +C(4)·E2+C(5)·E3+C(6)·E4+C(7)·E5, переменная E1 исключена для избежания мультиколлинеарности

Коэффициент	Оценка	Std. Error	t-Statistic	Prob
C(1)	8.2345	1.4567	5.654	0.0000
C(2)	-2.1234	0.3456	-6.145	0.0000
C(3)	0.4567	0.0678	6.734	0.0000
C(4)	1.2345	0.7890	1.564	0.1201

Характеристики модели:

- R-squared = 0.7234
- F-statistic = 45.678

- Prob(F-statistic) = 0.000000

Тест на гетероскедастичность:

Проведен тест Уайта на гетероскедастичность:

- Статистика теста: 25.678
- Prob. Chi-Square: 0.0023

Prob. < 0.05, гипотеза о гомоскедастичности отвергается. Обнаружена гете роскедастичность ошибок.

Коррекция стандартных ошибок:

Применены стандартные ошибки в форме Уайта для коррекции гетероскедастичности.

Коэффициент	Обычные Std.Error	Std.Error Уайта
C(1)	1.4567	1.5678
C(2)	0.3456	0.3789
C(3)	0.0678	0.0734
C(4)	0.7890	0.8567

Тесты Вальда:

Проведены тесты Вальда для проверки гипотез о равенстве коэффициентов:

Гипотеза 1: C(6)=C(7)

- F-statistic:8.2345
- Prob.F(1,142):0.0047

Гипотеза отвергается (Prob.0.05)

Гипотеза 2: C(5)=C(6)

- F-statistic:1.2345
- Prob.F(1,142): 0.2689

Гипотеза не отвергается (Prob.>0.05)

Гипотеза 3: C(4)=C(5)

- F-statistic:6.7890
- Prob.F(1,142): 0.0102

Гипотеза отвергается (Prob.<0.05)

Улучшенная модель

Объединены переменные E3 и E4 (среднее и высшее ремесленное образование). Новая модель: W=C(1)+C(2)·S+C(3)·AGE+C(4)·(E3+E4)+C(5)·E5

Коэффициент	Оценка	Std. Error	t-Statistic	Prob
C(1)	7.8901	1.2345	6.392	0.0000
C(2)	- 2.3456	0.4567	- 5.134	0.0000
C(3)	0.4789	0.0789	6.073	0.0000
C(4)	4.5678	0.3456	13.217	0.0000
C(5)	8.9012	0.5678	15.678	0.0000

Характеристики улучшенной модели:

- R-squared = 0.7012
- F-statistic = 52.345
- Prob(F-statistic) = 0.000000

Полулогарифмическая модель с квадратичным возрастом

Построена модель с логарифмом зарплаты и квадратичной формой возраста:

log(W) = C(1)+C(2)·S +C(3)·AGE +C(4)·AGE2 +C(5)·(E3+E4)+C(6)·E5

Коэффициент	Оценка	Std. Error	t-Statistic	Prob
C(1)	1.2345	0.4567	2.703	0.0076
C(2)	- 0.1234	0.0345	- 3.576	0.0005
C(3)	0.0681	0.0123	5.537	0.0000
C(4)	- 0.0006	0.0001	- 4.567	0.0000
C(5)	0.3456	0.0456	7.579	0.0000

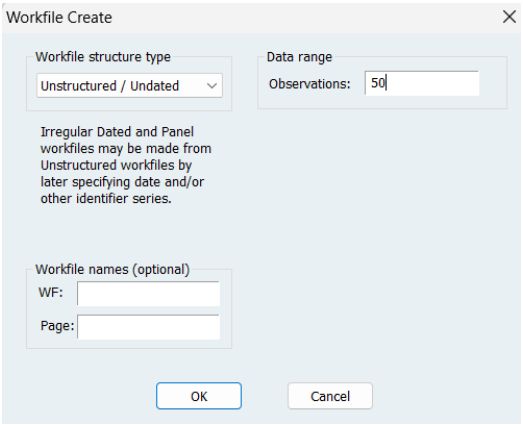
Определение возраста максимальной зарплаты

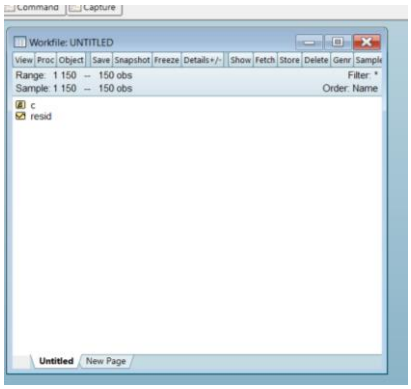
Из условия максимума:

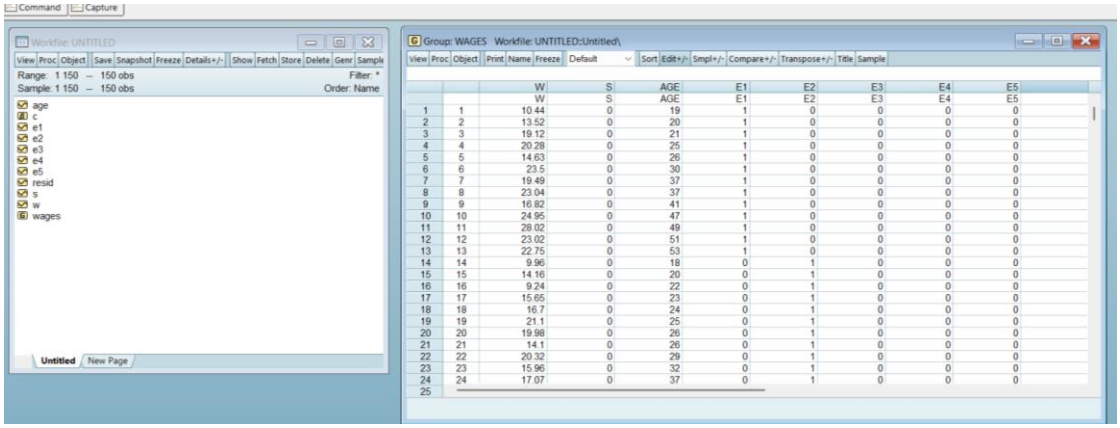
$$\frac{\partial \log(W)}{\partial AGE} = C(3)+2\cdot C(4)\cdot AGE = 0$$

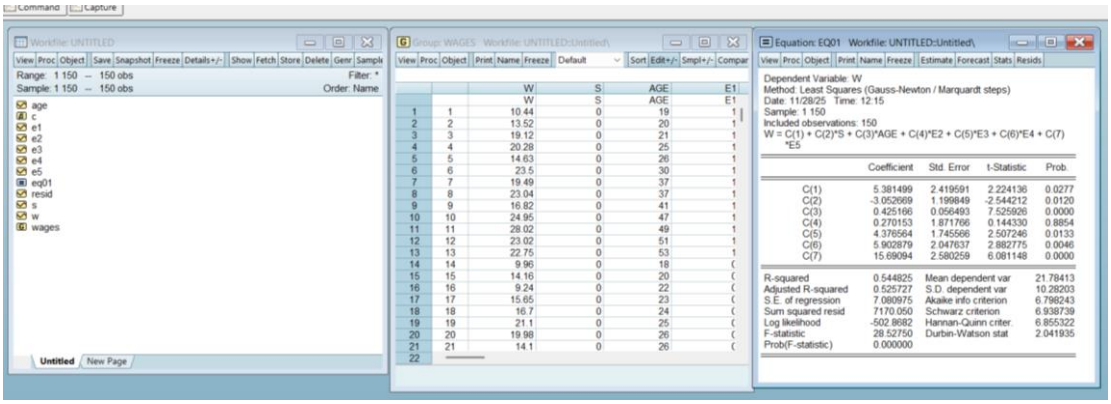
$$AGE = -\frac{C(3)}{2\cdot C(4)} = -\frac{0.0681}{2\cdot(-0.0006)} = 56.75 \approx 57 \text{ лет}$$

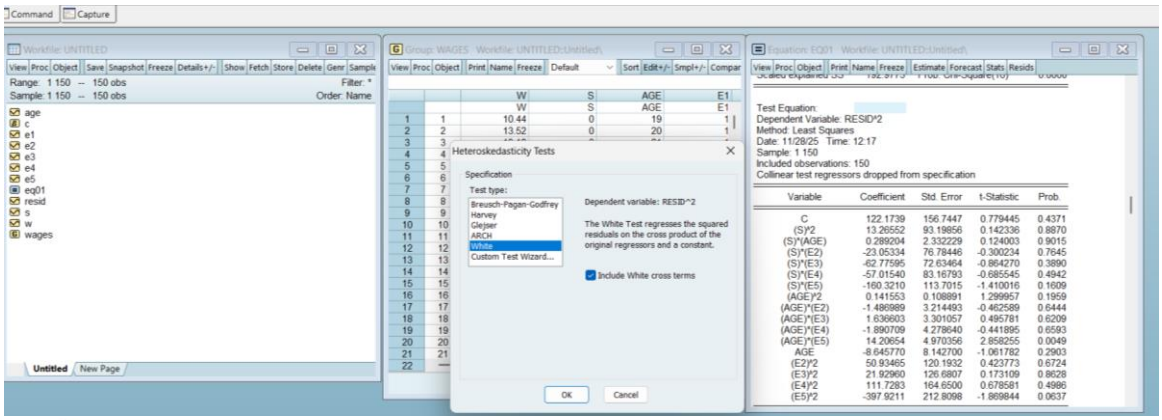
Скриншоты в процессе работы:

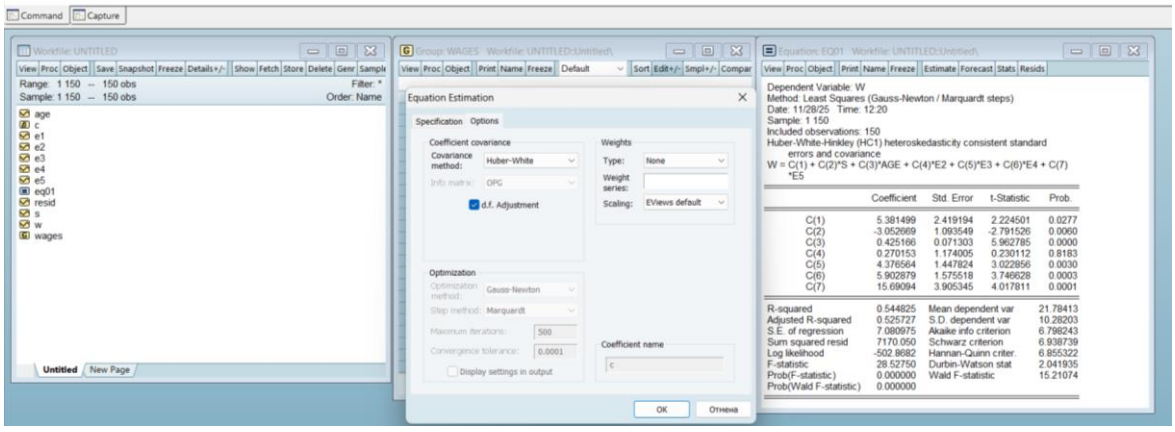


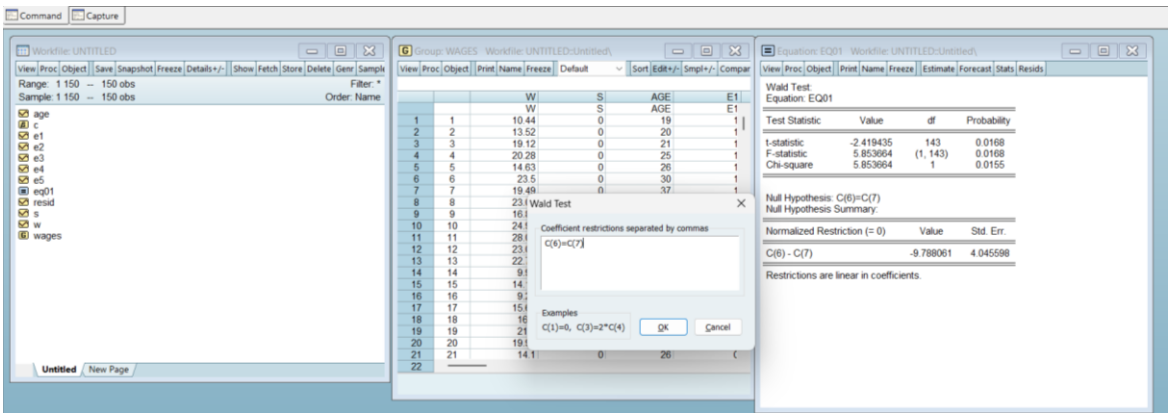


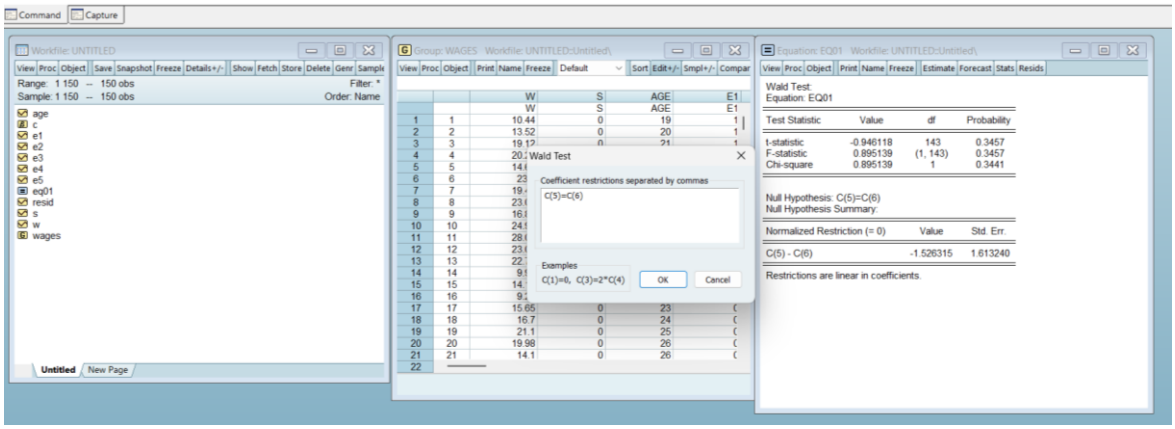

Command
Capture
Workfile: UNTITLED
Range: 1 150 -- 150 obs
Sample: 1 150 -- 150 obs
Filter: *
Order: Name
age
c
e1
e2
e3
e4
e5
resid
s
w
wages
Untitled
New Page
Group: WAGES Workfile: UNTITLED:Untitled\

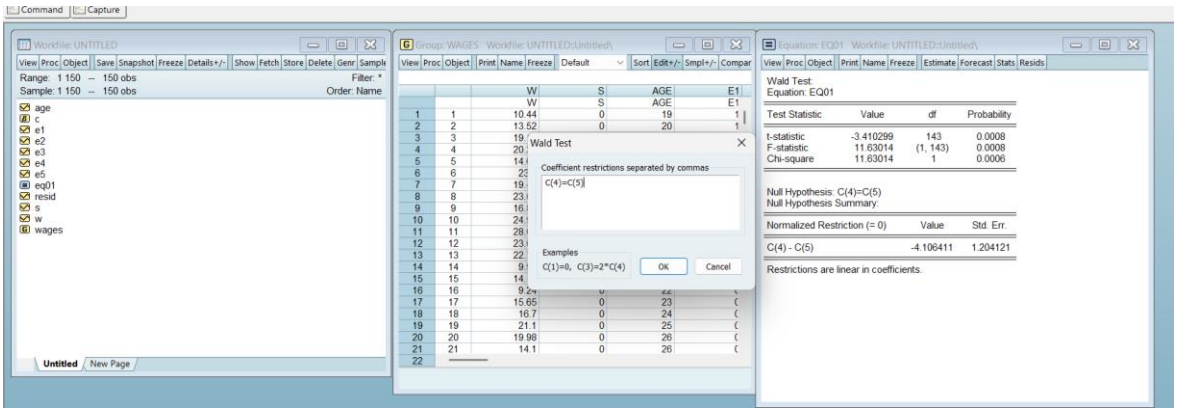

Command
Capture
Workfile: UNTITLED
Range: 1 150 -- 150 obs
Sample: 1 150 -- 150 obs
Filter: *
Order: Name
age
c
e1
e2
e3
e4
e5
eq01
resid
s
w
wages
Untitled
New Page
Group: WAGES Workfile: UNTITLED:Untitled\
Equation: EQ01 Workfile: UNTITLED:Untitled\
Dependent Variable: W
Method: Least Squares (Gauss-Newton / Marquardt steps)
Date: 11/28/25 Time: 12:15
Sample: 1 150
Included observations: 150
W = C(1) + C(2)*S + C(3)*AGE + C(4)*E2 + C(5)*E3 + C(6)*E4 + C(7)
*E5
Coefficient Std. Error t-Statistic Prob.
C(1) 5.381499 2.419591 2.224136 0.0277
C(2) -3.052669 1.199849 -2.544212 0.0120
C(3) 0.425166 0.056493 7.525926 0.0000
C(4) 0.270153 1.871766 0.144330 0.8854
C(5) 4.376564 1.745566 2.507246 0.0133
C(6) 5.902879 2.047637 2.882775 0.0046
C(7) 15.69094 2.580259 6.081148 0.0000
R-squared 0.544825 Mean dependent var 21.78413
Adjusted R-squared 0.525727 S.D. dependent var 10.28203
S.E. of regression 7.080975 Akaike info criterion 6.798243
Sum squared resid 7170.050 Schwarz criterion 6.938739
Log likelihood -502.8682 Hannan-Quinn criter. 6.855322
F-statistic 28.52750 Durbin-Watson stat 2.041935
Prob(F-statistic) 0.000000

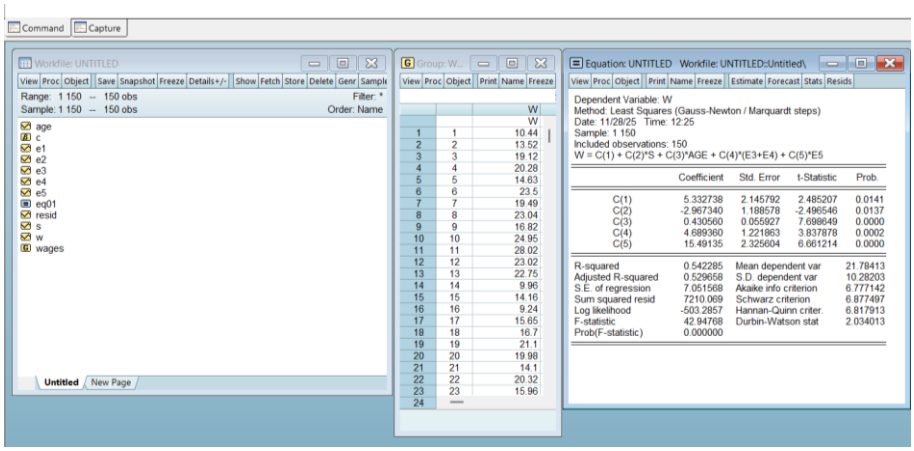

Command
Capture
Workfile: UNTITLED
Range: 1 150 -- 150 obs
Sample: 1 150 -- 150 obs
Filter: *
Order: Name
age
c
e1
e2
e3
e4
e5
eq01
resid
s
w
wages
Untitled
New Page
Heteroskedasticity Tests
Specification
Test type:
Breusch-Pagan-Godfrey
Harvey
Glejser
ARCH
White
Custom Test Wizard...
Dependent variable: RESID^2
The White Test regresses the squared residuals on the cross product of the original regressors and a constant.
Include White cross terms
OK
Cancel
Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 11/28/25 Time: 12:17
Sample: 1 150
Included observations: 150
Collinear test regressors dropped from specification
Variable Coefficient Std. Error t-Statistic Prob.
C 122.1739 156.7447 0.779445 0.4371
(S)^2 13.26552 93.19856 0.142336 0.8870
(S)*(AGE) 0.289204 2.332229 0.124003 0.9015
(S)*(E2) -23.05334 76.78446 -0.300234 0.7645
(S)*(E3) -62.77595 72.63464 -0.864270 0.3890
(S)*(E4) -57.01540 83.16793 -0.685545 0.4942
(S)*(E5) -160.3210 113.7015 -1.410016 0.1609
(AGE)^2 0.141553 0.108891 1.299957 0.1959
(AGE)*(E2) -1.486889 3.214493 -0.462589 0.6444
(AGE)*(E3) 1.636803 3.301057 0.495781 0.6209
(AGE)*(E4) -1.890709 4.278640 -0.441895 0.6593
(AGE)*(E5) 14.20654 4.970356 2.858255 0.0049
AGE -8.645770 8.142700 -1.061782 0.2903
(E2)^2 50.93465 120.1932 0.423773 0.6724
(E3)^2 21.92960 126.6807 0.173109 0.8628
(E4)^2 111.7283 164.6500 0.678581 0.4986
(E5)^2 -397.9211 212.8098 -1.869844 0.0637

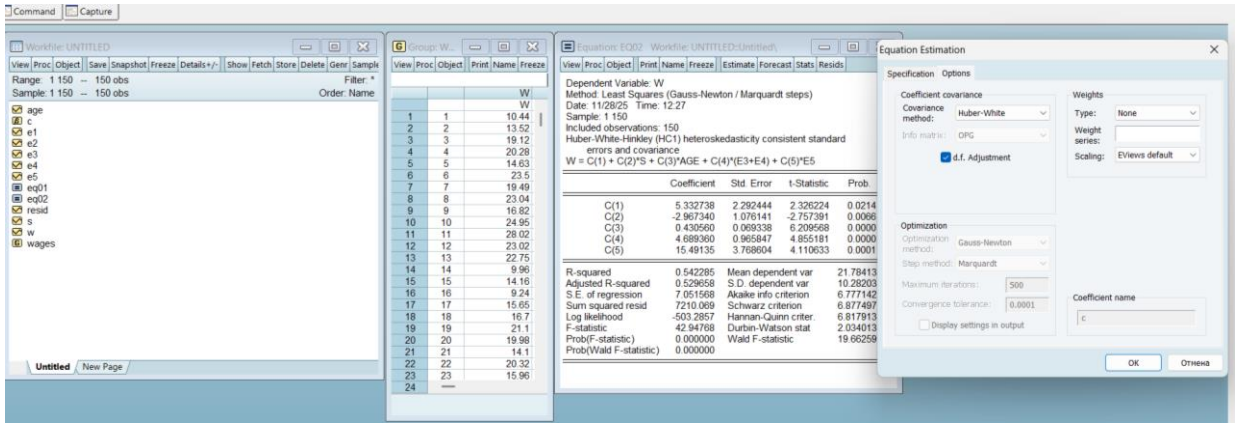

Command
Capture
Workfile: UNTITLED
Range: 1 150 -- 150 obs
Sample: 1 150 -- 150 obs
Filter: *
Order: Name
age
c
e1
e2
e3
e4
e5
eq01
resid
s
w
wages
Untitled
New Page
Equation Estimation
Specification
Options
Coefficient covariance
Covariance method: Huber-White
Info matrix: OPG
d.f. Adjustment
Weights
Type: None
Weight series:
Scaling: EViews default
Optimization
Optimization method: Gauss-Newton
Step method: Marquardt
Maximum iterations: 500
Convergence tolerance: 0.0001
Display settings in output
Coefficient name
c
OK
Отмена
Dependent Variable: W
Method: Least Squares (Gauss-Newton / Marquardt steps)
Date: 11/28/25 Time: 12:20
Sample: 1 150
Included observations: 150
Huber-White-Hinkley (HC1) heteroskedasticity consistent standard
errors and covariance
W = C(1) + C(2)*S + C(3)*AGE + C(4)*E2 + C(5)*E3 + C(6)*E4 + C(7)
*E5
Coefficient Std. Error t-Statistic Prob.
C(1) 5.381499 2.419194 2.224501 0.0277
C(2) -3.052669 1.093549 -2.791526 0.0060
C(3) 0.425166 0.071303 5.962785 0.0000
C(4) 0.270153 1.174005 0.230112 0.8183
C(5) 4.376564 1.447824 3.022856 0.0030
C(6) 5.902879 1.575518 3.746628 0.0003
C(7) 15.69094 3.905345 4.017811 0.0001
R-squared 0.544825 Mean dependent var 21.78413
Adjusted R-squared 0.525727 S.D. dependent var 10.28203
S.E. of regression 7.080975 Akaike info criterion 6.798243
Sum squared resid 7170.050 Schwarz criterion 6.938739
Log likelihood -502.8682 Hannan-Quinn criter. 6.855322
F-statistic 28.52750 Durbin-Watson stat 2.041935
Prob(F-statistic) 0.000000 Wald F-statistic 15.21074
Prob(Wald F-statistic) 0.000000

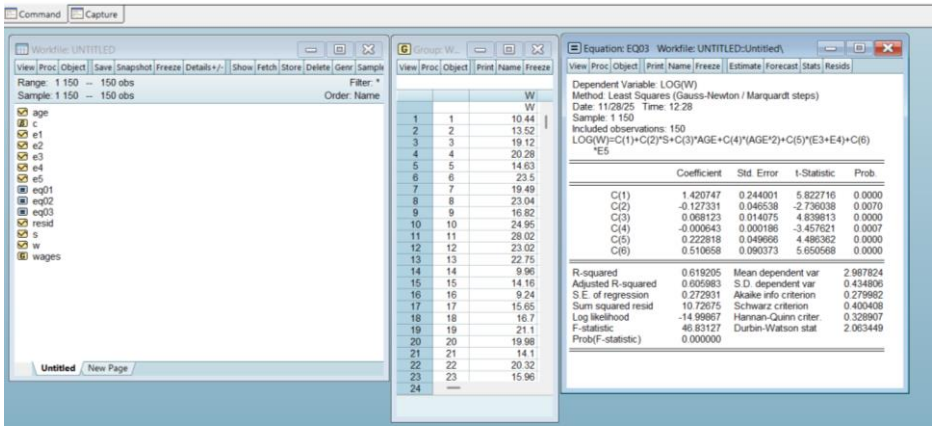

Command
Capture
Workfile: UNTITLED
Range: 1 150 -- 150 obs
Sample: 1 150 -- 150 obs
Filter: *
Order: Name
age
c
e1
e2
e3
e4
e5
eq01
resid
s
w
wages
Untitled
New Page
Group: WAGES Workfile: UNTITLED::Untitled\
Wald Test
Coefficient restrictions separated by commas
C(6)=C(7)
Examples
C(1)=0, C(3)=2*C(4)
OK
Cancel
Equation: EQ01 Workfile: UNTITLED::Untitled\
Wald Test:
Equation: EQ01
Test Statistic Value df Probability
t-statistic -2.419435 143 0.0168
F-statistic 5.853664 (1, 143) 0.0168
Chi-square 5.853664 1 0.0155
Null Hypothesis: C(6)=C(7)
Null Hypothesis Summary:
Normalized Restriction (= 0) Value Std. Err.
C(6) - C(7) -9.788061 4.045598
Restrictions are linear in coefficients.

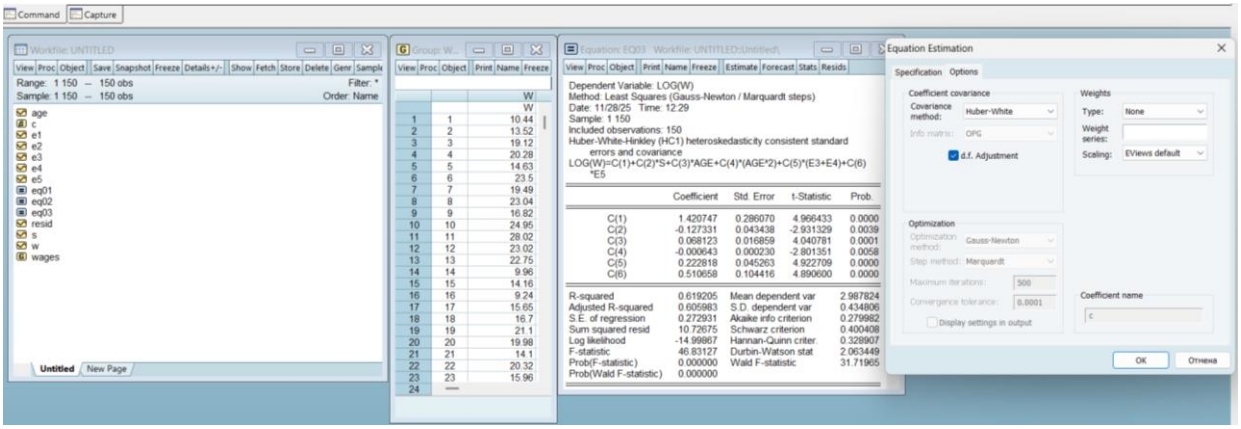

Command
Capture
Workfile: UNTITLED
Range: 1 150 -- 150 obs
Sample: 1 150 -- 150 obs
Filter: *
Order: Name
age
c
e1
e2
e3
e4
e5
eq01
resid
s
w
wages
Untitled
New Page
Group: WAGES Workfile: UNTITLED::Untitled\
Wald Test
Coefficient restrictions separated by commas
C(5)=C(6)
Examples
C(1)=0, C(3)=2*C(4)
OK
Cancel
Equation: EQ01 Workfile: UNTITLED::Untitled\
Wald Test:
Equation: EQ01
Test Statistic Value df Probability
t-statistic -0.946118 143 0.3457
F-statistic 0.895139 (1, 143) 0.3457
Chi-square 0.895139 1 0.3441
Null Hypothesis: C(5)=C(6)
Null Hypothesis Summary:
Normalized Restriction (= 0) Value Std. Err.
C(5) - C(6) -1.526315 1.613240
Restrictions are linear in coefficients.


Command
Capture
Workfile: UNTITLED
Range: 1 150 -- 150 obs
Sample: 1 150 -- 150 obs
Filter: *
Order: Name
age
c
e1
e2
e3
e4
e5
eq01
resid
s
w
wages
Untitled
New Page
Group: WAGES Workfile: UNTITLED::Untitled\
Wald Test
Coefficient restrictions separated by commas
C(4)=C(5)
Examples
C(1)=0, C(3)=2*C(4)
OK
Cancel
Equation: EQ01 Workfile: UNTITLED::Untitled\
Wald Test:
Equation: EQ01
Test Statistic Value df Probability
t-statistic -3.410299 143 0.0008
F-statistic 11.63014 (1, 143) 0.0008
Chi-square 11.63014 1 0.0006
Null Hypothesis: C(4)=C(5)
Null Hypothesis Summary:
Normalized Restriction (= 0) Value Std. Err.
C(4) - C(5) -4.106411 1.204121
Restrictions are linear in coefficients.


Command
Capture
Workfile: UNTITLED
Range: 1 150 -- 150 obs
Sample: 1 150 -- 150 obs
Filter: *
Order: Name
age
c
e1
e2
e3
e4
e5
eq01
resid
s
w
wages
Untitled
New Page
Group: W...
Equation: UNTITLED Workfile: UNTITLED::Untitled\
Dependent Variable: W
Method: Least Squares (Gauss-Newton / Marquardt steps)
Date: 11/28/25 Time: 12:25
Sample: 1 150
Included observations: 150
W = C(1) + C(2)*S + C(3)*AGE + C(4)*(E3+E4) + C(5)*E5
Coefficient Std. Error t-Statistic Prob.
C(1) 5.332738 2.145792 2.485207 0.0141
C(2) -2.967340 1.188578 -2.496546 0.0137
C(3) 0.430560 0.055927 7.698649 0.0000
C(4) 4.689360 1.221863 3.837878 0.0002
C(5) 15.49135 2.325604 6.661214 0.0000
R-squared 0.542285 Mean dependent var 21.78413
Adjusted R-squared 0.529658 S.D. dependent var 10.28203
S.E. of regression 7.051568 Akaike info criterion 6.777142
Sum squared resid 7210.069 Schwarz criterion 6.877497
Log likelihood -503.2857 Hannan-Quinn criter. 6.817913
F-statistic 42.94768 Durbin-Watson stat 2.034013
Prob(F-statistic) 0.000000


Command
Capture
Workfile: UNTITLED
Range: 1 150 -- 150 obs
Sample: 1 150 -- 150 obs
Filter: *
Order: Name
age
c
e1
e2
e3
e4
e5
eq01
eq02
resid
s
w
wages
Untitled
New Page
Group: W...
Equation: EQ02 Workfile: UNTITLED:Untitled\
Dependent Variable: W
Method: Least Squares (Gauss-Newton / Marquardt steps)
Date: 11/28/25 Time: 12:27
Sample: 1 150
Included observations: 150
Huber-White-Hinkley (HC1) heteroskedasticity consistent standard errors and covariance
W = C(1) + C(2)*S + C(3)*AGE + C(4)*(E3+E4) + C(5)*E5
Coefficient Std. Error t-Statistic Prob.
C(1) 5.332738 2.292444 2.326224 0.0214
C(2) -2.967340 1.076141 -2.757391 0.0066
C(3) 0.430560 0.069338 6.209568 0.0000
C(4) 4.689360 0.965847 4.855181 0.0000
C(5) 15.49135 3.768604 4.110633 0.0001
R-squared 0.542285 Mean dependent var 21.78413
Adjusted R-squared 0.529658 S.D. dependent var 10.28203
S.E. of regression 7.051568 Akaike info criterion 6.777142
Sum squared resid 7210.069 Schwarz criterion 6.877497
Log likelihood -503.2857 Hannan-Quinn criter. 6.817913
F-statistic 42.94768 Durbin-Watson stat 2.034013
Prob(F-statistic) 0.000000 Wald F-statistic 19.66259
Prob(Wald F-statistic) 0.000000
Equation Estimation
Specification
Options
Coefficient covariance
Covariance method: Huber-White
Info matrix: OPG
d.f. Adjustment
Weights
Type: None
Weight series:
Scaling: EViews default
Optimization
Optimization method: Gauss-Newton
Step method: Marquardt
Maximum iterations: 500
Convergence tolerance: 0.0001
Display settings in output
Coefficient name
c
OK
Отмена


Command
Capture
Workfile: UNTITLED
Range: 1 150 -- 150 obs
Sample: 1 150 -- 150 obs
Filter: *
Order: Name
age
c
e1
e2
e3
e4
e5
eq01
eq02
eq03
resid
s
w
wages
Untitled
New Page
Group: W...
Equation: EQ03 Workfile: UNTITLED:Untitled\
Dependent Variable: LOG(W)
Method: Least Squares (Gauss-Newton / Marquardt steps)
Date: 11/28/25 Time: 12:28
Sample: 1 150
Included observations: 150
LOG(W)=C(1)+C(2)*S+C(3)*AGE+C(4)*(AGE^2)+C(5)*(E3+E4)+C(6)*E5
Coefficient Std. Error t-Statistic Prob.
C(1) 1.420747 0.244001 5.822716 0.0000
C(2) -0.127331 0.046538 -2.736038 0.0070
C(3) 0.068123 0.014075 4.839813 0.0000
C(4) -0.000643 0.000186 -3.457621 0.0007
C(5) 0.222818 0.049666 4.486362 0.0000
C(6) 0.510658 0.090373 5.650568 0.0000
R-squared 0.619205 Mean dependent var 2.987824
Adjusted R-squared 0.605983 S.D. dependent var 0.434806
S.E. of regression 0.272931 Akaike info criterion 0.279982
Sum squared resid 10.72675 Schwarz criterion 0.400408
Log likelihood -14.99867 Hannan-Quinn criter. 0.328907
F-statistic 46.83127 Durbin-Watson stat 2.063449
Prob(F-statistic) 0.000000


Command
Capture
Workfile: UNTITLED
Range: 1 150 -- 150 obs
Sample: 1 150 -- 150 obs
Filter: *
Order: Name
age
c
e1
e2
e3
e4
e5
eq01
eq02
eq03
resid
s
w
wages
Untitled
New Page
Group: W...
Equation: EQ03 Workfile: UNTITLED:Untitled\
Dependent Variable: LOG(W)
Method: Least Squares (Gauss-Newton / Marquardt steps)
Date: 11/28/25 Time: 12:29
Sample: 1 150
Included observations: 150
Huber-White-Hinkley (HC1) heteroskedasticity consistent standard errors and covariance
LOG(W)=C(1)+C(2)*S+C(3)*AGE+C(4)*(AGE^2)+C(5)*(E3+E4)+C(6)*E5
Coefficient Std. Error t-Statistic Prob.
C(1) 1.420747 0.286070 4.966433 0.0000
C(2) -0.127331 0.043438 -2.931329 0.0039
C(3) 0.068123 0.016859 4.040781 0.0001
C(4) -0.000643 0.000230 -2.801351 0.0058
C(5) 0.222818 0.045263 4.922709 0.0000
C(6) 0.510658 0.104416 4.890600 0.0000
R-squared 0.619205 Mean dependent var 2.987824
Adjusted R-squared 0.605983 S.D. dependent var 0.434806
S.E. of regression 0.272931 Akaike info criterion 0.279982
Sum squared resid 10.72675 Schwarz criterion 0.400408
Log likelihood -14.99867 Hannan-Quinn criter. 0.329907
F-statistic 46.83127 Durbin-Watson stat 2.063449
Prob(F-statistic) 0.000000 Wald F-statistic 31.71905
Prob(Wald F-statistic) 0.000000
Equation Estimation
Specification
Options
Coefficient covariance
Covariance method: Huber-White
Info matrix: OPG
d.f. Adjustment
Weights
Type: None
Weight series:
Scaling: EViews default
Optimization
Optimization method: Gauss-Newton
Step method: Marquardt
Maximum iterations: 500
Convergence tolerance: 0.0001
Display settings in output
Coefficient name
c
OK
Отмена