

La régression linéaire Multiple

Par DR Djellab



La régression linéaire Multiple

- Méthode statistique.
- Modélisation de la relation (une variable dépendante (ou réponse) et + variables indépendantes (ou prédicteurs)).

Le modèle de régression linéaire multiple s'exprime par l'équation suivante :

$$Y = c + a_1X_1 + a_2X_2 + \dots + a_nX_n + \epsilon$$

Telque

a_1x_1 et a_2x_2 sont les coefficients de régression qui indique combien Y change lorsque X_1 et X_2 augmentent d'une unité.



La régression linéaire Multiple

Objectif:

Etudier l'impact de l'expérience et l'éducation sur le salaire.

$$\text{Salaire} = c + a_1 * \text{experience} + a_2 * \text{education}$$



EViews

La régression linéaire Multiple

The screenshot shows the EViews main window with the 'Quick' menu open. The menu items are: Sample..., Generate Series..., Show..., Graph..., Empty Group (Edit Series), Series Statistics, Group Statistics, Estimate Equation..., and Estimate VAR... The 'Estimate Equation...' option is highlighted. The main window displays a list of variables on the left, including 'wage', 'educ', and 'exper', and a menu bar at the top with 'File', 'Edit', 'Object', 'View', 'Proc', 'Quick', 'Options', 'Window', and 'Help'.

The 'Equation Estimation' dialog box is shown with the 'Specification' tab selected. The 'Equation specification' section contains the text 'wage c educ exper'. The 'Estimation settings' section shows 'Method: LS - Least Squares (NLS and ARMA)' and 'Sample: 1 526'. The dialog box has 'OK' and 'Annuler' buttons at the bottom right.



La régression linéaire Multiple

$$\text{Salaire} = c + a1 * \text{experience} + a2 * \text{education}$$

$$\text{Salaire} = -3,39 + 0,07\text{experience} + 0,64\text{education}.$$

Equation: UNTITLED Workfile: UNTITLED::Untitled\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: WAGE
Method: Least Squares
Date: 11/13/24 Time: 13:43
Sample: 1 526
Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.390539	0.766566	-4.423023	0.0000
EXPER	0.070095	0.010978	6.385291	0.0000
EDUC	0.644272	0.053806	11.97397	0.0000

R-squared	0.225162	Mean dependent var	5.896103
Adjusted R-squared	0.222199	S.D. dependent var	3.693086
S.E. of regression	3.257044	Akaike info criterion	5.205204
Sum squared resid	5548.160	Schwarz criterion	5.229531
Log likelihood	-1365.969	Hannan-Quinn criter.	5.214729
F-statistic	75.98998	Durbin-Watson stat	1.820274
Prob(F-statistic)	0.000000		

Std.error: écart type, Précision des coefficients estimés (plus c'est bas, mieux c'est).

t-statistic: Test de significativité pour chaque coefficient (valeur absolue élevée = significatif).

F-statistic : Significativité globale du modèle (plus élevé = modèle globalement significatif).

R-squared : Ajustement du modèle (plus c'est proche de 1 mieux le modèle



La régression linéaire
Multiple

$$\text{Salaire} = c + a1 * \text{expe} + a2 * \text{education}$$

$$\text{Salaire} = -3,39 + 0,07\text{expe} + 0,64\text{education}.$$



Hypothèses liées à la regression linéaire multiple:

1. La normalité des résidus.
2. L'Homoscédasticité.
3. Absence de multicollinéarité.



Test de Normalité

Si le modèle respecte l'hypothèse de la normalité des erreurs:

H_0 = les erreurs suivent une loi normal.

Test de jarque bira

The screenshot shows the EViews software interface. The main window title is "Equation: UNTITLED Workfile: UNTITLED::Untitled\". The menu bar includes "View", "Proc", "Object", "Print", "Name", "Freeze", "Estimate", "Forecast", "Stats", and "Resids". The "Resids" menu is open, showing options like "Representations", "Estimation Output", "Coefficient Labels", "Actual,Fitted,Residual", "ARMA Structure...", "Gradients and Derivatives", "Covariance Matrix", "Coefficient Diagnostics", "Residual Diagnostics", "Stability Diagnostics", and "Label". The "Residual Diagnostics" option is selected, and a sub-menu is displayed with options: "Correlogram - Q-statistics...", "Correlogram Squared Residuals...", "Histogram - Normality Test", "Serial Correlation LM Test...", and "Heteroskedasticity Tests...". The "Histogram - Normality Test" option is highlighted. Below the menu, a table displays the following data:

	Std. Error	t-Statistic	Prob.
	0.766566	-4.423023	0.0000
	0.053806	11.97397	0.0000
	0.010978	6.385291	0.0000

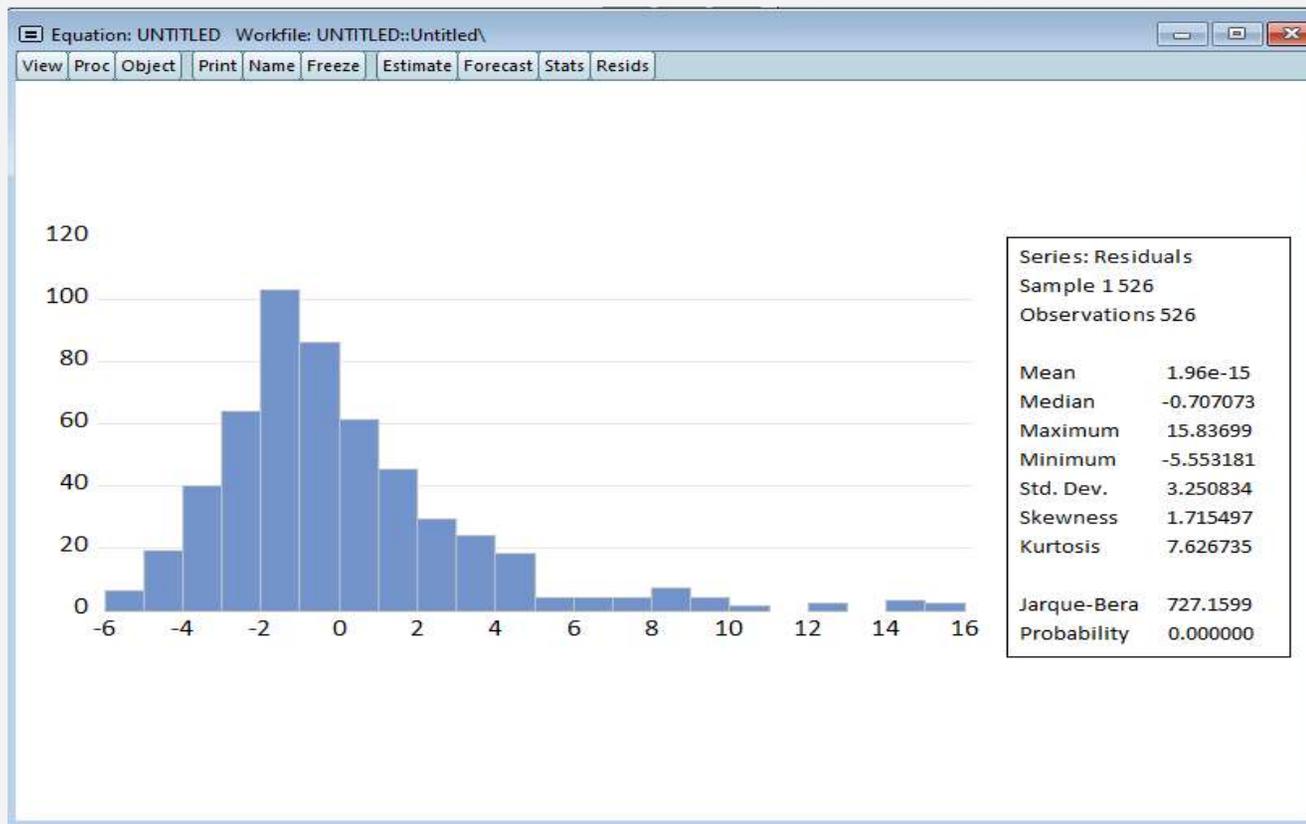
At the bottom of the window, the following statistics are shown:

F-statistic	75.98998
Prob(F-statistic)	0.000000



EViews

Test de Normalité





homoscédasticité

H0: « il y'a une homoscédasticité des erreurs »

On veut prouvé que les erreurs ont une variances constantes.

The screenshot shows the EViews software interface. At the top, it says "Equation: UNTITLED Workfile: UNTITLED::Untitled\". Below this is a menu bar with options: View, Proc, Object, Print, Name, Freeze, Estimate, Forecast, Stats, Resids. A dropdown menu is open under "View", listing various options: Representations, Estimation Output, Coefficient Labels, Actual, Fitted, Residual, ARMA Structure..., Gradients and Derivatives, Covariance Matrix, Coefficient Diagnostics, Residual Diagnostics (highlighted), Stability Diagnostics, and Label. To the right of the menu, a table displays regression statistics:

	Std. Error	t-Statistic	Prob.
	0.766566	-4.423023	0.0000
	0.053806	11.97397	0.0000
	0.010978	6.385291	0.0000

At the bottom of the window, a summary table shows:

F-statistic	75.98998
Prob(F-statistic)	0.000000

A secondary dropdown menu is open under "Residual Diagnostics", listing several tests: Correlogram - Q-statistics..., Correlogram Squared Residuals..., Histogram - Normality Test, Serial Correlation LM Test..., and Heteroskedasticity Tests... (highlighted).



EViews

Homoscédasticité

Equation: UNTITLED Workfile: UNTITLED::Untitled\

View	Proc	Object	Print	Name	Freeze	Estimate	Forecast	Stats	Resids
------	------	--------	-------	------	--------	----------	----------	-------	--------

Heteroskedasticity Test: Breusch-Pagan-Godfrey
Null hypothesis: Homoskedasticity

F-statistic	16.85322	Prob. F(2,523)	0.0000
Obs*R-squared	31.84728	Prob. Chi-Square(2)	0.0000
Scaled explained SS	104.3215	Prob. Chi-Square(2)	0.0000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 11/14/24 Time: 10:37
Sample: 1 526
Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-23.23086	6.211818	-3.739783	0.0002
EDUC	2.163007	0.436014	4.960867	0.0000
EXPER	0.388162	0.088957	4.363499	0.0000

R-squared	0.060546	Mean dependent var	10.54783
Adjusted R-squared	0.056954	S.D. dependent var	27.17855
S.E. of regression	26.39324	Akaike info criterion	9.389780
Sum squared resid	364323.5	Schwarz criterion	9.414107
Log likelihood	-2466.512	Hannan-Quinn criter.	9.399305
F-statistic	16.85322	Durbin-Watson stat	1.960873
Prob(F-statistic)	0.000000		

Heteroskedasticity Tests

Specification

Test type:

- Breusch-Pagan-Godfrey**
- Harvey
- Glejser
- ARCH
- White
- Custom Test Wizard...

Dependent variable: RESID^2

The Breusch-Pagan-Godfrey Test regresses the squared residuals on the original regressors by default.

Regressors:

c educ exper

Add equation regressors

OK Cancel



EViews

homoscédasticité

Equation: UNTITLED Workfile: UNTITLED::Untitled\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic	11.93737	Prob. F(5,520)	0.0000
Obs*R-squared	54.15905	Prob. Chi-Square(5)	0.0000
Scaled explained SS	177.4077	Prob. Chi-Square(5)	0.0000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 11/14/24 Time: 10:41
Sample: 1 526
Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	51.70462	21.28490	2.429169	0.0155
EDUC^2	0.381434	0.098866	3.858087	0.0001
EDUC*EXPER	0.102413	0.036520	2.804310	0.0052
EDUC	-9.093897	2.860470	-3.179162	0.0016
EXPER^2	-0.014881	0.007185	-2.071211	0.0388
EXPER	-0.219108	0.622333	-0.352075	0.7249

R-squared	0.102964	Mean dependent var	10.54783
Adjusted R-squared	0.094339	S.D. dependent var	27.17855
S.E. of regression	25.86480	Akaike info criterion	9.354984
Sum squared resid	347873.8	Schwarz criterion	9.403638
Log likelihood	-2454.361	Hannan-Quinn criter.	9.374035
F-statistic	11.93737	Durbin-Watson stat	1.983759
Prob(F-statistic)	0.000000		

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



EViews

Detection de la multi colinéarité

The screenshot shows the EViews software interface. The 'Quick' menu is open, and 'Group Statistics' is selected. A sub-menu is visible, showing 'Correlations' as the selected option. Below the menu, a table displays coefficients for variables C, EDUC, and EXPER.

variable	Coefficient
C	-3.39053
EDUC	0.64427
EXPER	0.07009

Below the table, the R-squared value is 0.225162 and the Mean dependent variable is 5.896103.

The 'Series List' dialog box is shown, containing a list of series and/or series expressions. The list contains 'educ' and 'exper'. The dialog has 'OK' and 'Cancel' buttons at the bottom.



EViews

Detection de la multi colinéarité

Group: UNTITLED Workfile: U...

View	Proc	Object	Print	Name	Freeze	Sample	Sheet
Correlation							
		EDUC		EXPER			
EDUC		1.000000		-0.299542			^
EXPER		-0.299542		1.000000			
							^
							^
							^
							^



Detection de la multi colinéarité

Variance inflation factor

Equation: UNTITLED Workfile: UNTITLED::Untitled\

View	Proc	Object	Print	Name	Freeze	Estimate	Forecast	Stats	Resids
Variance Inflation Factors									
Date: 11/14/24 Time: 10:56									
Sample: 1 526									
Included observations: 526									
Variable	Coefficient Variance	Uncentered VIF	Centered VIF						
C	0.587624	29.13652	NA						
EDUC	0.002895	23.75386	1.098569						
EXPER	0.000121	2.828895	1.098569						

Equation: UNTITLED Workfile: UNTITLED::Untitled\

View	Proc	Object	Print	Name	Freeze	Estimate	Forecast	Stats	Resids
Representations									
Estimation Output									
Coefficient Labels									
Actual, Fitted, Residual									
ARMA Structure...									
Gradients and Derivatives									
Covariance Matrix									
						Std. Error	t-Statistic	Prob.	
						0.766566	-4.423023	0.0000	
						0.053806	11.97397	0.0000	
Coefficient Diagnostics									
Residual Diagnostics									
Stability Diagnostics									
Label									
F-statistic					75.98998				
Prob(F-statistic)					0.000000				

- Scaled Coefficients
- Confidence Intervals...
- Confidence Ellipse...
- Variance Inflation Factors
- Coefficient Variance Decomposition
- Wald Test- Coefficient Restrictions...
- Omitted Variables Test - Likelihood Ratio...
- Redundant Variables Test - Likelihood Ratio...
- Factor Breakpoint Test...



Correction des problèmes de régression

- ✓ L'hétéroscédasticité dans un modèle de régression multiple \Rightarrow la variance des erreurs n'est pas constante.
- ✓ Viole de l'hypothèse fondamentale
- ✓ Des estimations inefficaces et des intervalles de confiance incorrects.



Les méthodes utilisées pour remédier
au problème de l'hétéroscédasticité

Correction des problèmes de régression

Transformation des variables en logarithmes.

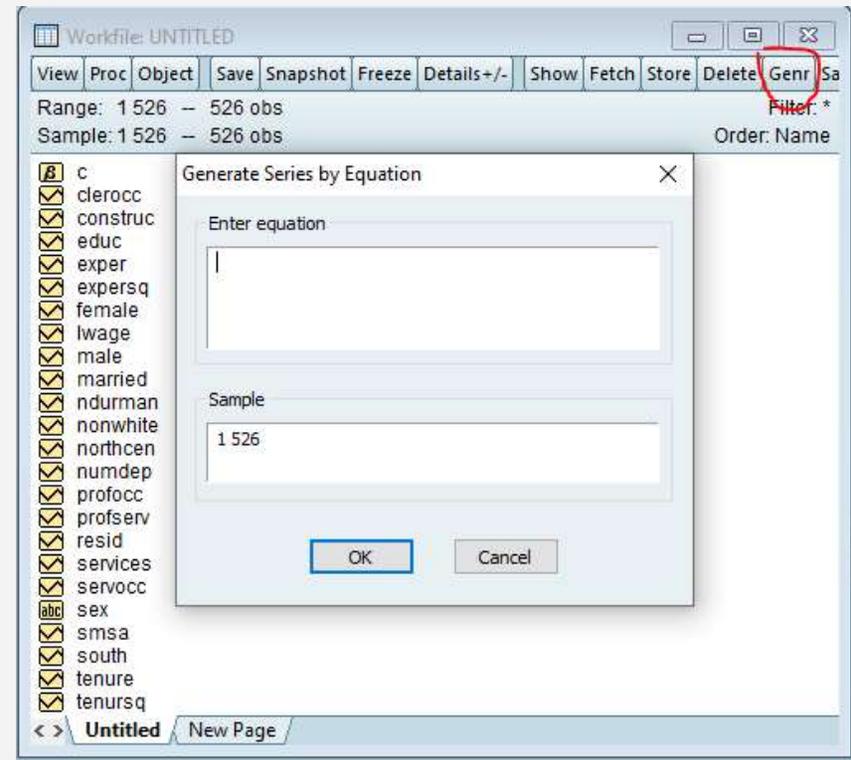
Méthode de variance covariance.



Correction des problèmes de régression

Transformation des variables

- Log(wage)
- Log(educ)
- Log(exper)





EViews

Correction des problèmes de régression

Generate Series by Equation

Enter equation

$\log_wage = \log(wage)$

Sample

1 526

OK Cancel

Generate Series by Equation

Enter equation

$\log_educ = \log(educ)$

Sample

1 526

OK Cancel

Generate Series by Equation

Enter equation

$\log_exper = \log(exper)$

Sample

1 526

OK Cancel



Correction des problèmes de régression

Ré estimation du modèle à nouveau

Equation Estimation

Specification Options

Equation specification

Dependent variable followed by list of regressors including ARMA and PDL terms, OR an explicit equation like $Y=c(1)+c(2)^*X$.

log_wage c log_educ log_exper|

Estimation settings

Method: LS - Least Squares (NLS and ARMA)

Sample: 1 526

OK Annuler

Equation: UNTITLED Workfile: UNTITLED::Untitled\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: LOG_WAGE
Method: Least Squares
Date: 12/05/24 Time: 09:22
Sample: 1 526
Included observations: 524

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.229470	0.223007	-5.513157	0.0000
LOG_EXPER	0.167513	0.019310	8.675074	0.0000
LOG_EDUC	0.978471	0.082794	11.81818	0.0000

R-squared	0.256073	Mean dependent var	1.624714
Adjusted R-squared	0.253217	S.D. dependent var	0.531916
S.E. of regression	0.459663	Akaike info criterion	1.289064
Sum squared resid	110.0823	Schwarz criterion	1.313462
Log likelihood	-334.7347	Hannan-Quinn criter.	1.298618
F-statistic	89.66889	Durbin-Watson stat	1.761659
Prob(F-statistic)	0.000000		



Correction des problèmes de régression

Test d'homoscédasticité

Equation: UNTITLED Workfile: UNTITLED::Untitled\

	Std. Error	t-Statistic	Prob.
	0.223007	-5.513157	0.0000
	0.019310	8.675074	0.0000
	0.082794	11.81818	0.0000

Residual Diagnostics

- Correlogram - Q-statistics...
- Correlogram Squared Residuals...
- Histogram - Normality Test
- Serial Correlation LM Test...
- Heteroskedasticity Tests...

F-statistic 89.66889
Prob(F-statistic) 0.000000

Heteroskedasticity Tests

Specification

Test type: Breusch-Pagan-Godfrey

Dependent variable: RESID^2

The Breusch-Pagan-Godfrey Test regresses the squared residuals on the original regressors by default.

Regressors: c log_exper log_educ

Add equation regressors

OK Cancel



EViews

Correction des problèmes de régression

Equation: UNTITLED Workfile: UNTITLED::Untitled

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Heteroskedasticity Test: Breusch-Pagan-Godfrey
Null hypothesis: Homoskedasticity

F-statistic	8.124635	Prob. F(2,521)	0.0003
Obs*R-squared	15.84854	Prob. Chi-Square(2)	0.0004
Scaled explained SS	20.09940	Prob. Chi-Square(2)	0.0000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 12/05/24 Time: 09:27
Sample: 1 526
Included observations: 524

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.014292	0.161230	-0.088643	0.9294
LOG_EXPER	0.056155	0.013961	4.022386	0.0001
LOG_EDUC	0.035941	0.059859	0.600437	0.5485

R-squared	0.030245	Mean dependent var	0.210081
Adjusted R-squared	0.026523	S.D. dependent var	0.336827
S.E. of regression	0.332330	Akaike info criterion	0.640331
Sum squared resid	57.54088	Schwarz criterion	0.664729
Log likelihood	-164.7668	Hannan-Quinn criter.	0.649886
F-statistic	8.124635	Durbin-Watson stat	2.074089

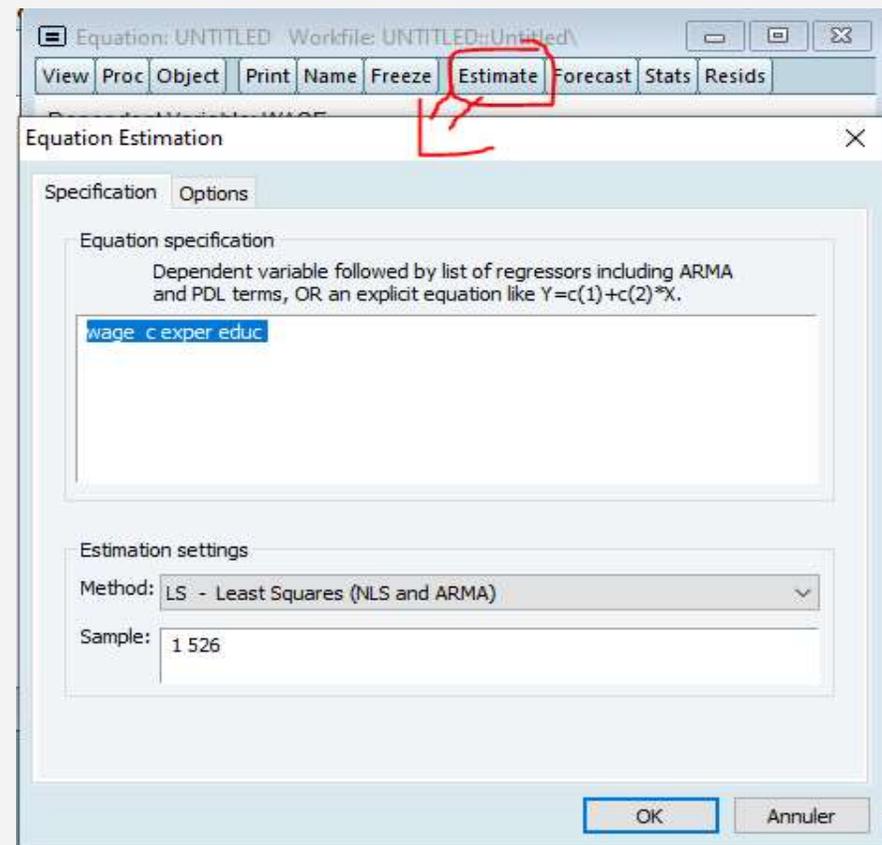
P-value < 0,05, on rejette h₀

Problème non corrigé.



Correction des problèmes de régression

2 eme méthode : matrice de variance covariance





Correction des problèmes de régression

2 eme méthode : matrice de variance covariance

Equation Estimation ✕

Specification Options

Coefficient covariance

Covariance method: Ordinary ▾

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

Outliers / indicator saturation

Auto-detection Options

Coefficient name

OK Annuler



Correction des problèmes de régression

2 eme méthode : matrice de variance covariance

Equation Estimation

Specification Options

Coefficient covariance

Covariance method: Ordinary

Info matrix: HC (various)

Huber-White
HAC (Newey-West)
Cluster-robust

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

Outliers / indicator saturation

Auto-detection Options

Coefficient name

c

OK Annuler



EViews

Correction des problèmes de régression

2 eme méthode : matrice de variance covariance

Equation: UNTITLED Workfile: UNTITLED::Untitled\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: WAGE
Method: Least Squares
Date: 12/05/24 Time: 09:45
Sample: 1 526
Included observations: 526
Huber-White-Hinkley (HC1) heteroskedasticity consistent standard errors and covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.390539	0.864875	-3.920267	0.0001
EXPER	0.070095	0.010994	6.375622	0.0000
EDUC	0.644272	0.065187	9.883457	0.0000

R-squared	0.225162	Mean dependent var	5.896103
Adjusted R-squared	0.222199	S.D. dependent var	3.693086
S.E. of regression	3.257044	Akaike info criterion	5.205204
Sum squared resid	5548.160	Schwarz criterion	5.229531
Log likelihood	-1365.969	Hannan-Quinn criter.	5.214729
F-statistic	75.98998	Durbin-Watson stat	1.820274
Prob(F-statistic)	0.000000	Wald F-statistic	50.31634
Prob(Wald F-statistic)	0.000000		

Table: UNTITLED Workfile: UNTITLED::Untitled\

View Proc Object Print Name Edit+/- CellFmt Grid+/- Title Comments+/-

	A	B	C	D	E
1	Dependent Variable: WAGE				
2	Method: Least Squares				
3	Date: 12/05/24 Time: 09:32				
4	Sample: 1 526				
5	Included observations: 526				
6					
7	Variable	Coefficient	Std. Error	t-Statistic	Prob.
8					
9	C	-3.390539	0.766566	-4.423023	0.0000
10	EXPER	0.070095	0.010978	6.385291	0.0000
11	EDUC	0.644272	0.053806	11.97397	0.0000
12					
13	R-squared	0.225162	Mean dependent var	5.896103	
14	Adjusted R-squared	0.222199	S.D. dependent var	3.693086	
15	S.E. of regression	3.257044	Akaike info criterion	5.205204	
16	Sum squared resid	5548.160	Schwarz criterion	5.229531	
17	Log likelihood	-1365.969	Hannan-Quinn criter.	5.214729	
18	F-statistic	75.98998	Durbin-Watson stat	1.820274	
19	Prob(F-statistic)	0.000000			
20					
21					
22					
23					
24					
25					