

Το τελευταίο βήμα της ανάλυσης παλινδρόμησης ...

ΔΙΑΓΝΩΣΤΙΚΟΙ ΕΛΕΓΧΟΙ :

- 1. Κανονικότητα $\rightarrow \varepsilon_i \sim N(0, \sigma^2)$

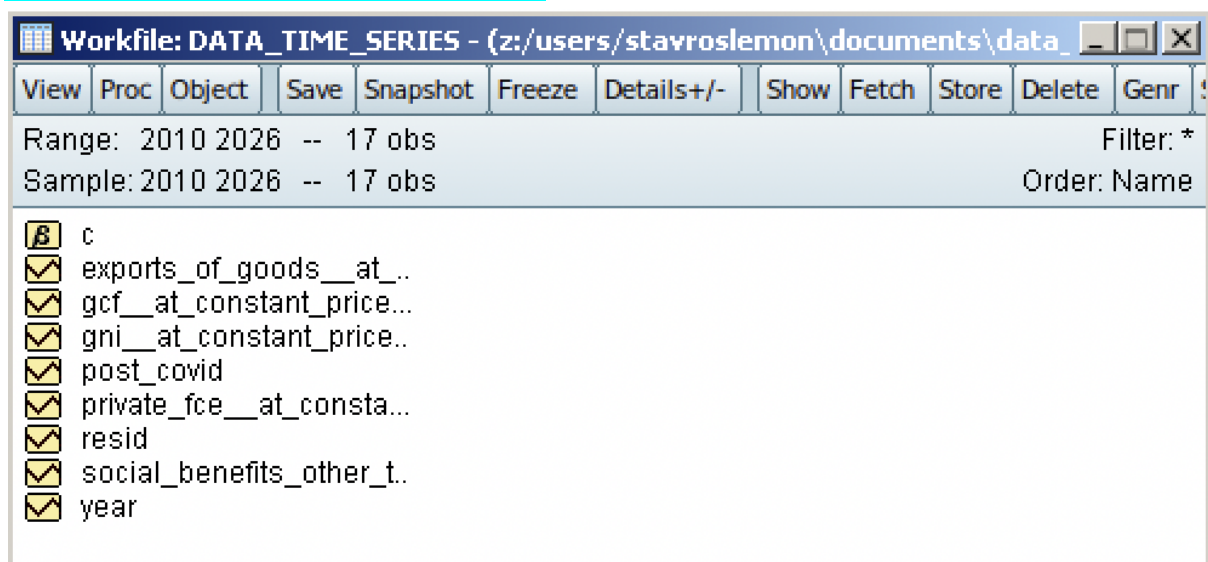
Με πιο απλά λόγια, ΔΕΝ θέλουμε να παραβιάζεται η υπόθεση της ΚΑΝΟΝΙΚΟΤΗΤΑΣ.

Βήμα 1ο: Τρέξτε την παλινδρόμηση.

Βήμα 2ο: Υπολογίστε και αποθηκεύστε τα κατάλοιπα ($residuals = Y_i - \hat{Y}_i$).

Βήμα 3ο: Χτίστε το διάγραμμα συχνοτήτων (βλ. ιστόγραμμα) των καταλοίπων.

Παράδειγμα χρονολογικών δεδομένων (time-series):



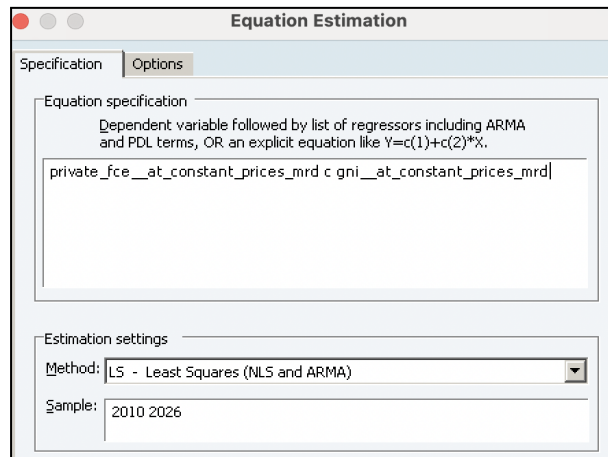
Ψ (εξαρτημένη): *private final consumption expenditure* = κατανάλωση

X_1 (ανεξάρτητη): *gross national income* = εισόδημα

$t = 2010 - 2026$, $i = \text{Greece}$

πηγή: Ευρωπαϊκή Επιτροπή (AMECO online)

Τρέχουμε στο enviews μια απλή γραμμική παλινδρόμηση (k=1):



Dependent Variable: PRIVATE_FCE__AT_CONSTANT_PRICES_MRD

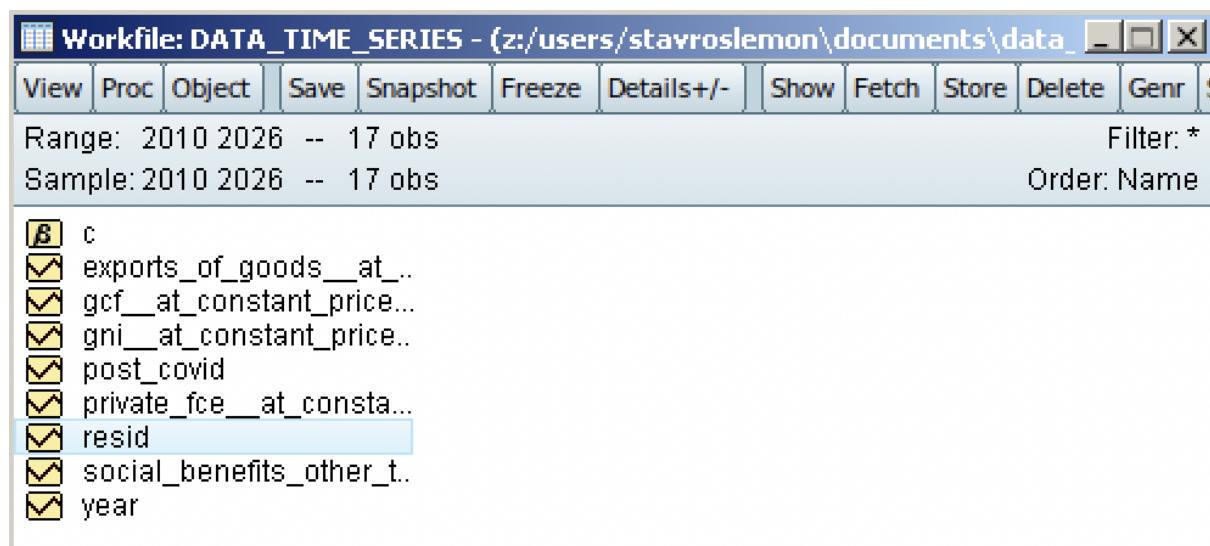
Method: Least Squares

Date: 05/05/25 Time: 15:58

Sample: 2010 2026

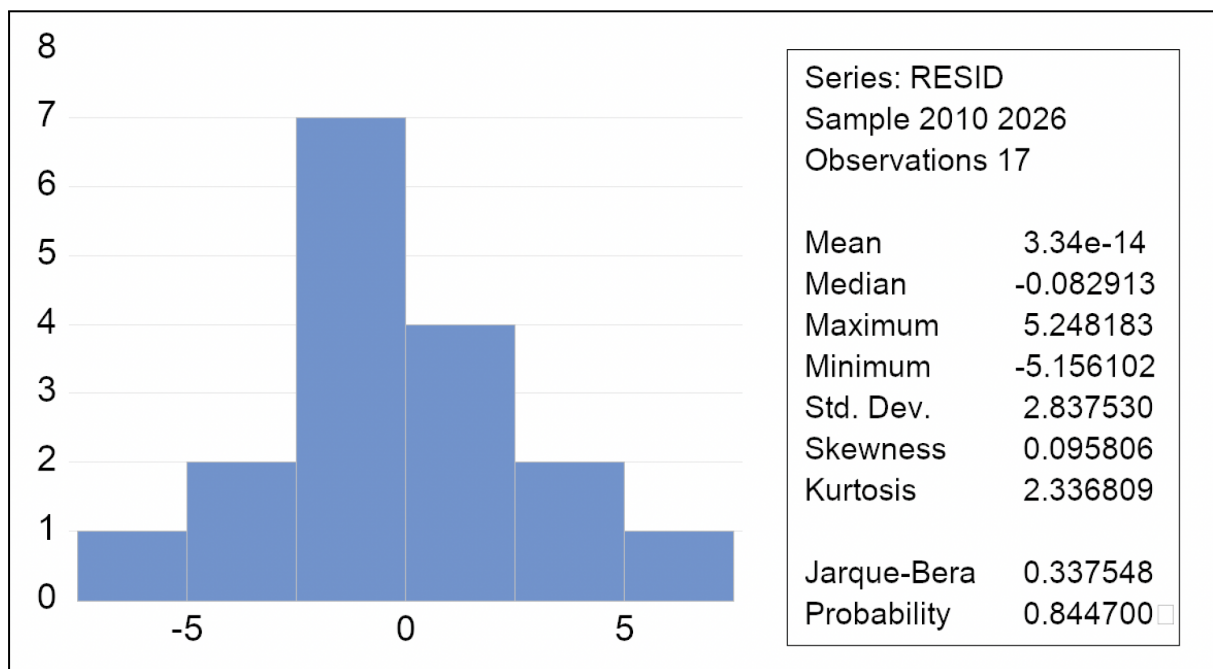
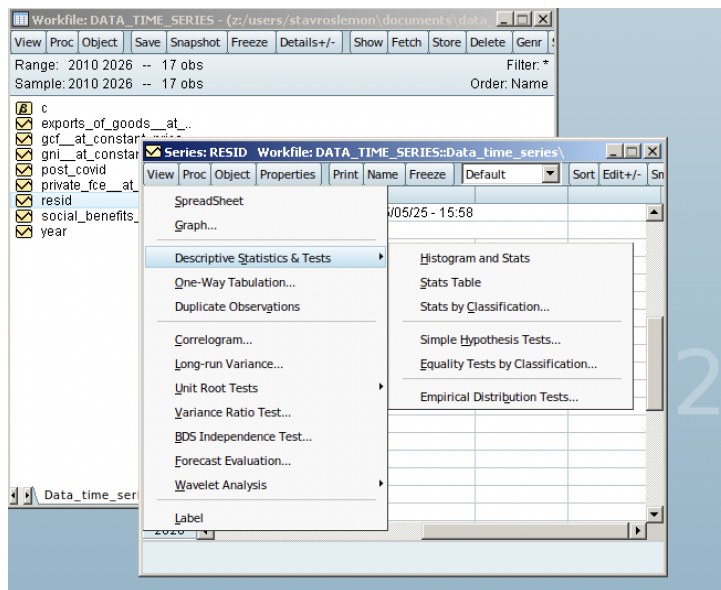
Included observations: 17

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-21.40112	10.65696	-2.008182	0.0630
GNI__AT_CONSTANT_PRICES_MRD	0.821767	0.057442	14.30598	0.0000
R-squared				
		0.931713	Mean dependent var	130.7176
Adjusted R-squared		0.927160	S.D. dependent var	10.85854
S.E. of regression		2.930588	Akaike info criterion	5.098414
Sum squared resid		128.8252	Schwarz criterion	5.196439
Log likelihood		-41.33652	Hannan-Quinn criter.	5.108158
F-statistic		204.6610	Durbin-Watson stat	1.204631
Prob(F-statistic)		0.000000		



Τα **κατάλοιπα** αποθηκεύονται αυτόματα όταν τρέχουμε μια παλινδρόμηση (βλ. **resid**)

Πρώτη Επιλογή: ανοίξετε τα κατάλοιπα και ζητήστε να σας δώσει το ιστόγραμμα ...



Εδώ βλέπετε το **ιστόγραμμα**, δηλαδή το διάγραμμα συχνοτήτων, των καταλοίπων. Η ένδειξη αριστερά δείχνει μια κατανομή που **φαίνεται να είναι κανονική** (μοιάζει με καμπάνα). **Θυμηθείτε**: μια κανονική κατανομή πρέπει να είναι **συμμετρική** → δηλ πρέπει να έχει συντελεστή ασυμμετρίας (**skewness**) = 0. Επιπλέον, μια κανονική κατανομή πρέπει να είναι **μεσόκυρτη** → δηλ πρέπει να έχει συντελεστή κύρτωσης (**kurtosis**) = 3.

Αν παρατηρήσετε τις τιμές του πίνακα των περιγραφικών στατιστικών ο **συντελεστής ασυμμετρίας** είναι **κοντά στο μηδέν** → **0,095** και ο **συντελεστής κύρτωσης** είναι **σχετικά κοντά στο 3** → **2,336**.

Την ακριβή απάντηση όμως θα την δώσει ο **έλεγχος των Jarque - Bera**:

- Διατύπωση των υποθέσεων του ελέγχου:

H₀: τα κατάλοιπα ακολουθούν την κανονική κατανομή

H₁: τα κατάλοιπα **ΔΕΝ** ακολουθούν την κανονική κατανομή

- Υπολογισμός της στατιστικής τιμής ελέγχου → **JB = 0,33**
(συνοδεύεται πάντα από την αντίστοιχη τιμή του **p-value = 0,84**)
- Σύγκριση της στατιστικής τιμής με την κριτική τιμή της κατανομής *Chi-Squared* (εξαγωγή συμπεράσματος)

Στην περίπτωση αυτή, χρησιμοποιούμε τον βασικό κανόνα για τα p-values:

Αν **p-value < 0,05 (5%)** → απορρίπτουμε την μηδενική υπόθεση

Αν **p-value > 0,05 (5%)** → ΔΕΝ μπορούμε να απορρίψουμε την μηδενική υπόθεση

Εμείς, έχουμε **πολύ μεγάλη του p-value = 0,84 > 0,05 (5%)** → επομένως,

ΔΕΝ μπορούμε να απορρίψουμε την μηδενική υπόθεση της κανονικότητας των καταλοίπων.

Δεύτερη Επιλογή: αφού τρέξετε την παλινδρόμηση → **view** → **residuals diagnostics** → **histogram normality test** (θα δώσει ακριβώς το ίδιο αποτέλεσμα)

Χωρίς τη βοήθεια του *enviews* η διαδικασία με το χέρι παίρνει λίγο χρόνο:

Στατιστική τιμή ελέγχου των J-B:

$$JB = \frac{n}{6} \left(S^2 + \frac{1}{4} (K - 3)^2 \right)$$

όπου:

$$S = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{3/2}}$$

(συντελεστής ασυμμετρίας)

$$K = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^2}$$

(συντελεστής κυρτότητας)

Τέλος, πρέπει να συγκρίνω την στατιστική τιμή ελέγχου με την **κριτική τιμή** από τους πίνακες της κατανομής χ^2 .

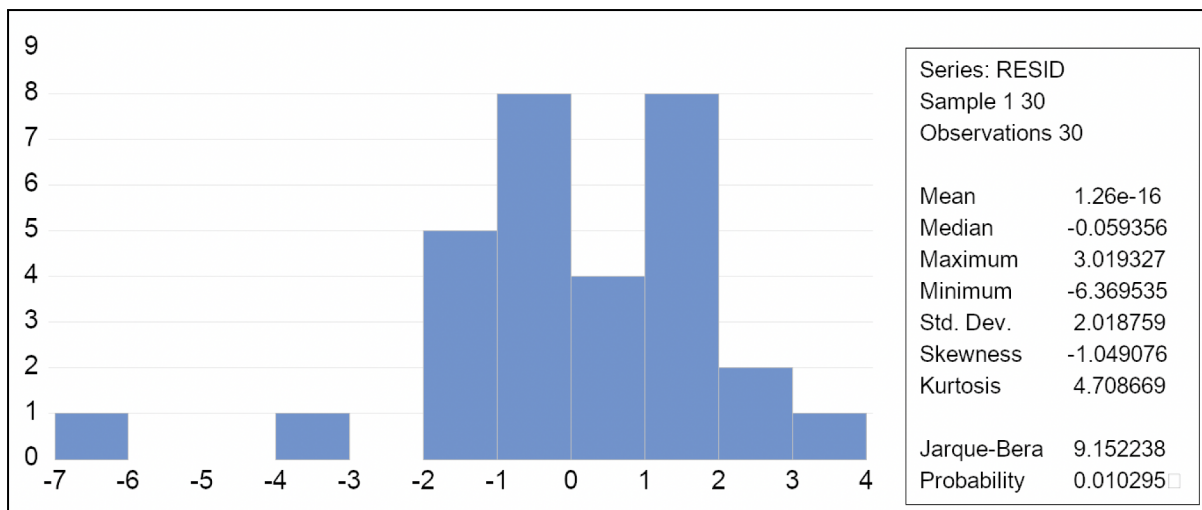
<https://www.numberanalytics.com/blog/mastering-jarque-bera-test-data-normality-guide>
<https://www.statology.org/how-to-read-chi-square-distribution-table/>

Παράδειγμα διαστρωματικών δεδομένων (cross-sectional data):

Εδώ, έχουμε στοιχεία από ένα γκρουπ 30 χωρών ($i=1,2,3,4,\dots,29,30$) κατά το έτος 2023 ($t=2023$). Παίρνουμε ως εξαρτημένη μεταβλητή (Ψ) τους ρυθμούς οικονομικής ανάπτυξης και ως κύρια ανεξάρτητη ($X1$) τον χρόνο που απαιτείται προκειμένου να ανοίξει μια επιχείρηση. **Δείτε τι θα συμβεί:**

Dependent Variable: GDP_GROWTH__ANNUAL____
Method: Least Squares
Date: 05/05/25 Time: 16:42
Sample: 1 30
Included observations: 30

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.976542	0.639239	1.527664	0.1378
TIME_REQUIRED_TO_START_A_BUSIN...	-0.012455	0.046423	-0.268298	0.7904
R-squared	0.002564	Mean dependent var		0.837667
Adjusted R-squared	-0.033058	S.D. dependent var		2.021353
S.E. of regression	2.054492	Akaike info criterion		4.342275
Sum squared resid	118.1863	Schwarz criterion		4.435688
Log likelihood	-63.13413	Hannan-Quinn criter.		4.372159
F-statistic	0.071984	Durbin-Watson stat		1.559870
Prob(F-statistic)	0.790437			



Παρατηρήστε τα αποτελέσματα: το ιστόγραμμα αριστερά δείχνει πως η **κορυφή αυτής της κατανομής ΔΕΝ είναι στη μέση**. Πολλές παρατηρήσεις φαίνεται να συγκεντρώνονται προς τα δεξιά. **Αρα, δε φαίνεται να είναι κανονική**. Ο **συντελεστής ασυμμετρίας** πρέπει να είναι κοντά στο μηδέν (εδώ είναι = **-1,04**). Ο **συντελεστής κύρτωσης** πρέπει να είναι κοντά στο 3 (εδώ είναι = **4,70**). Απάντηση με ακρίβεια θα δώσει ο **έλεγχος της κανονικότητας των Jarque - Bera**:

H₀: τα κατάλοιπα ακολουθούν την κανονική κατανομή

H₁: τα κατάλοιπα **ΔΕΝ** ακολουθούν την κανονική κατανομή

Στατιστική τιμή ελέγχου → **JB = 9,15** (αυτή τη φορά είναι πολύ μεγάλη ...)

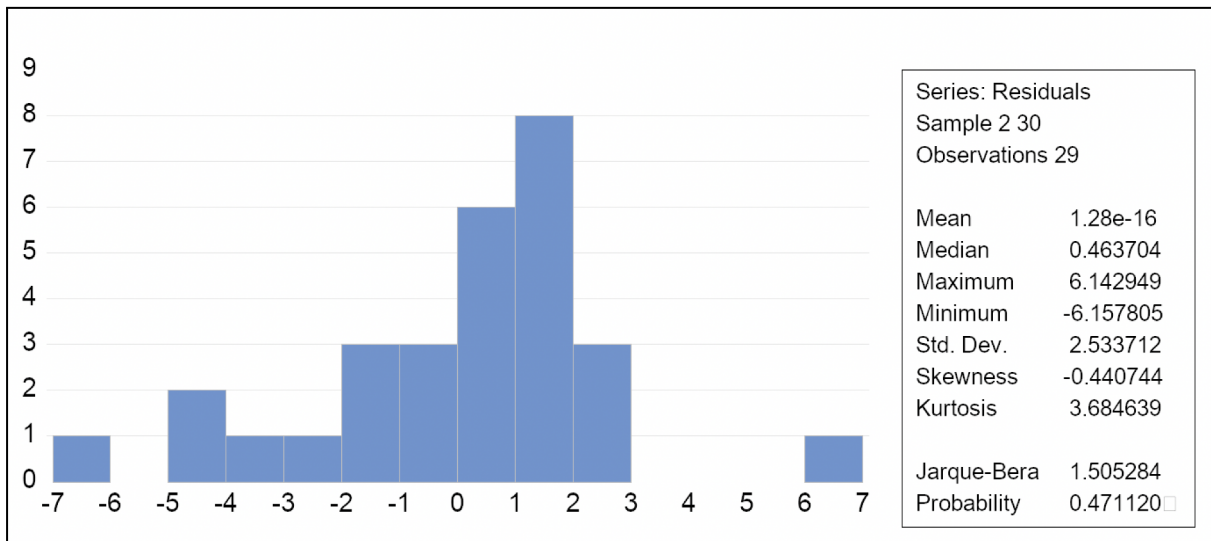
Το p-value που την συνοδεύει είναι πολύ μικρό → **p-value = 0,01 < 0,05 (5%)**.

Αυτό σημαίνει πως **απορρίπτω την μηδενική υπόθεση της κανονικότητας των καταλοίπων**. Αυτό, συνεπάγεται την **παραβίαση** μιας εκ των βασικών υποθέσεων του υποδείγματος. **Συνέπειες**: Τι συμβαίνει τώρα όταν παραβιάζεται η υπόθεση της κανονικότητας; **ΔΕΝ** επηρεάζονται οι εκτιμητές της παλινδρόμησης → **Όμως** επηρεάζονται τα **τυπικά σφάλματα** των εκτιμητών → και κατά συνέπεια οι **έλεγχοι υπόθεσης** για την σημαντικότητα των εκτιμητών, καθώς και τα **διαστήματα εμπιστοσύνης**.

Πως μπορώ να θεραπεύσω το πρόβλημα της παραβίασης της υπόθεσης της κανονικότητας;;;

Τροποποίηση του μοντέλου σας → πάρτε τις **πρώτες διαφορές** ή **λογαριθμίστε** ...

Χρήση των πρώτων διαφορών:



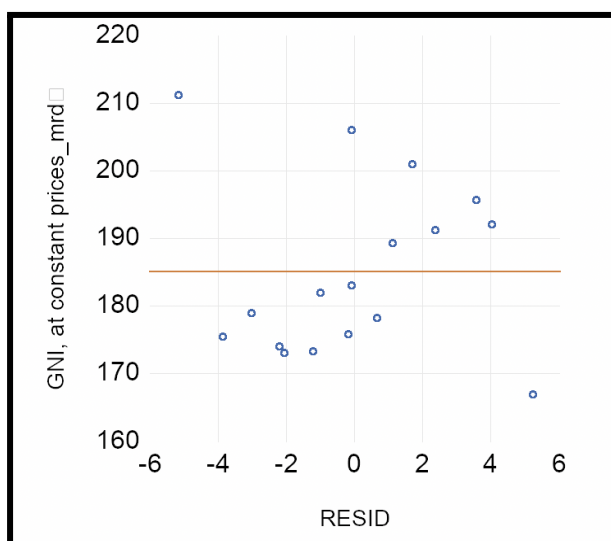
Τώρα, η τιμή του ελέγχου JB είναι σχετικά μικρή (**JB=1,50**), ενώ η αντίστοιχη τιμή του p-value είναι πολύ μεγάλη (**p-value=0,47 > 0,05**). Άρα, **ΔΕΝ** είμαστε σε θέση να απορρίψουμε την μηδενική υπόθεση της κανονικότητας των καταλοίπων (θεραπεύσαμε την παραβίαση της υπόθεσης της κανονικότητας ...).

2. Ομοσκεδαστικότητα:

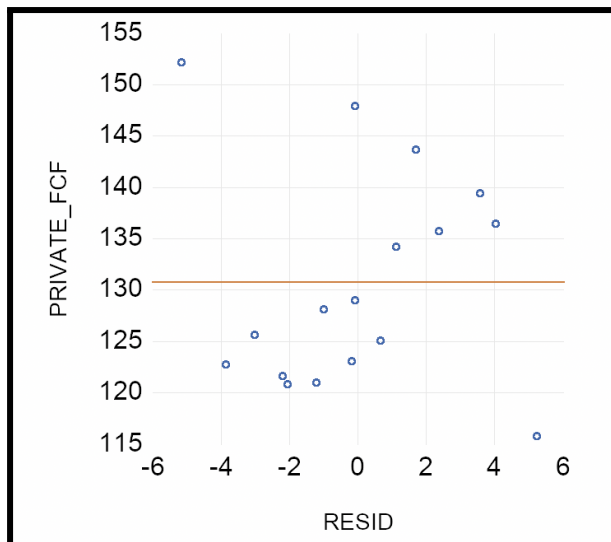
$$\varepsilon_i \sim N(0, \sigma^2) \rightarrow \text{διακύμανση σταθερή (ίση με } \sigma^2)$$

Γραφικές Ενδείξεις: κάντε το διάγραμμα διασποράς που να συνδέει τα **κατάλοιπα** με τις τιμές της **ανεξάρτητης μεταβλητής X** ή το αντίστοιχο που να συνδέει τα **κατάλοιπα** με τις **εκτιμώμενες τιμές της Ψ**. Κάθε μοτίβο που θυμίζει “χωνί” **ΔΕΝ** είναι καλή ένδειξη.

Επιστρέφουμε στο αρχικό μας μοντέλο χρονοσειρών:

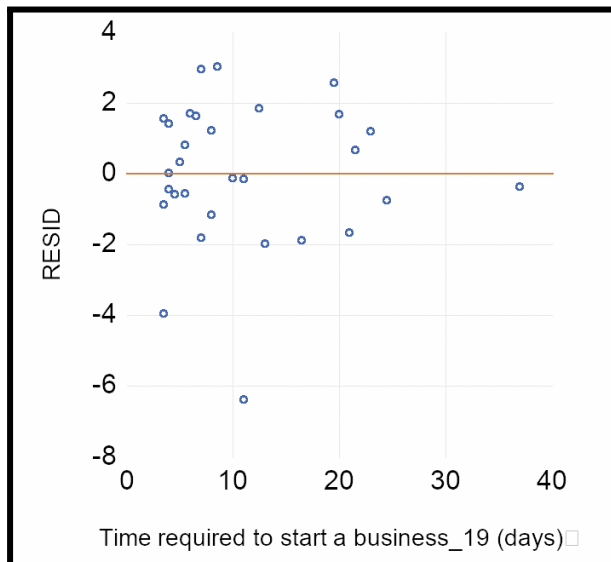


Διάγραμμα διασποράς: **κατάλοιπα και τιμές ανεξάρτητης μεταβλητής (X)**

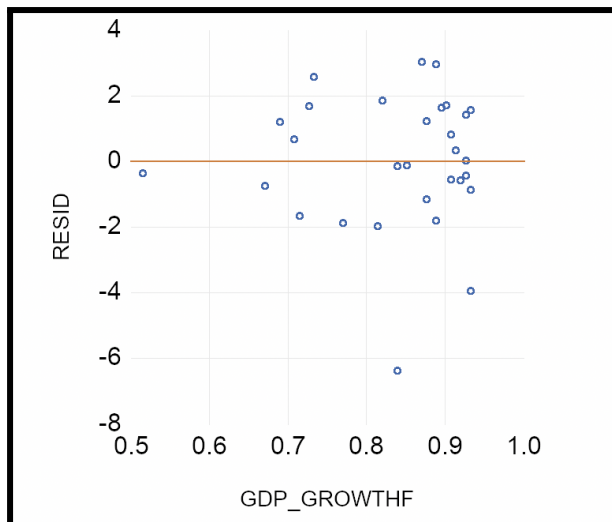


Διάγραμμα διασποράς: **κατάλοιπα και προβλεπόμενες τιμές της ψ**

Αντίστοιχα, η εικόνα στο μοντέλο με τα διαστρωματικά δεδομένα:

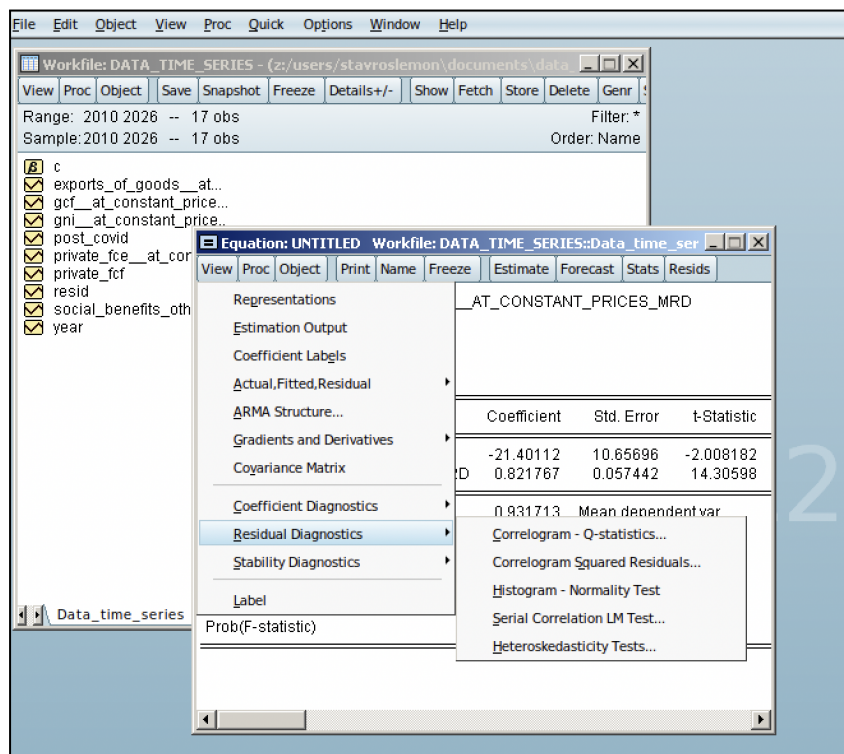


Διάγραμμα διασποράς: **κατάλοιπα και τιμές ανεξάρτητης μεταβλητής (X)**



Διάγραμμα διασποράς: **κατάλοιπα και προβλεπόμενες τιμές της ψ**

Στατιστικός Έλεγχος για την ύπαρξη **Ετεροσκεδαστικότητας**:



Το enviews δίνει **πολλές διαφορετικές επιλογές** ελέγχων για την ύπαρξη **ετεροσκεδαστικότητας** (βλ. παραβίαση της υπόθεσης της ομοσκεδαστικότητας)

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 gni_at_constant_prices_mrd

Add equation

OK Cancel

Για παράδειγμα, εδώ φαίνονται τα αποτελέσματα του ελέγχου των **Breusch - Pagan - Godfrey**:

Heteroskedasticity Test: Breusch-Pagan-Godfrey				
Null hypothesis: Homoskedasticity				
F-statistic	0.181306	Prob. F(1,15)	0.6763	
Obs*R-squared	0.203026	Prob. Chi-Square(1)	0.6523	
Scaled explained SS	0.105652	Prob. Chi-Square(1)	0.7452	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 05/05/25 Time: 17:40				
Sample: 2010 2026				
Included observations: 17				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-6.746340	33.71592	-0.200094	0.8441
GNI__AT_CONSTANT_PRICES_MRD	0.077382	0.181733	0.425801	0.6763
R-squared	0.011943	Mean dependent var	7.577953	
Adjusted R-squared	-0.053928	S.D. dependent var	9.031315	
S.E. of regression	9.271637	Akaike info criterion	7.401928	
Sum squared resid	1289.449	Schwarz criterion	7.499953	
Log likelihood	-60.91639	Hannan-Quinn criter.	7.411672	
F-statistic	0.181306	Durbin-Watson stat	2.113687	
Prob(F-statistic)	0.676300			

Η μηδενική υπόθεση του ελέγχου δέχεται την ύπαρξη **ομοσκεδαστικότητας**.
 Ακολουθεί η στατιστική τιμή του ελέγχου → εδώ, ακολουθεί την κατανομή F → **F=0,18**
 Η τιμή της στατιστικής F συνοδεύεται όπως πάντα από το αντίστοιχο p-value → εδώ **p-value=0,6763** → κανόνας: μεγάλη τιμή του p-value σημαίνει **ΜΗ** απόρριψη της μηδενικής υπόθεσης !!! Επομένως, επειδή **p-value = 0,67 > 0,05 (5%)** → **ΔΕΝ** μπορούμε να απορρίψουμε την μηδενική υπόθεση της ομοσκεδαστικότητας.

Με τον ίδιο ακριβώς τρόπο, **ΔΕΝ** μπορούμε να απορρίψουμε την μηδενική υπόθεση της ομοσκεδαστικότητας ούτε στο μοντέλο των διαστρωματικών δεδομένων:

Heteroskedasticity Test: Breusch-Pagan-Godfrey				
Null hypothesis: Homoskedasticity				
F-statistic	0.107254	Prob. F(1,28)	0.7457	
Obs*R-squared	0.114477	Prob. Chi-Square(1)	0.7351	
Scaled explained SS	0.184918	Prob. Chi-Square(1)	0.6672	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 05/05/25 Time: 17:51				
Sample: 1 30				
Included observations: 30				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.586266	2.438737	1.880591	0.0705
TIME_REQUIRED_TO_START_A_BUSIN...	-0.058002	0.177107	-0.327497	0.7457
R-squared	0.003816	Mean dependent var	3.939543	
Adjusted R-squared	-0.031762	S.D. dependent var	7.716432	
S.E. of regression	7.838020	Akaike info criterion	7.020190	
Sum squared resid	1720.167	Schwarz criterion	7.113603	
Log likelihood	-103.3028	Hannan-Quinn criter.	7.050074	
F-statistic	0.107254	Durbin-Watson stat	2.126954	
Prob(F-statistic)	0.745728			

Τι είναι όμως αυτό που ελέγχει ο έλεγχος για την ύπαρξη ετεροσκεδαστικότητας των Breusch - Pagan - Godfrey;;;

Αν υπάρχει κάποια γραμμική σχέση που να συνδέει τα τετράγωνα των καταλοίπων με τις τιμές των ανεξάρτητων μεταβλητών (X).

Ρίξτε μια ματιά στα αποτελέσματα στους παραπάνω πίνακες και δείτε τι παλινδρομεί.

Δείτε το αποτέλεσμα του ελέγχου όταν προσθέτουμε μια ακόμη ανεξάρτητη μεταβλητή στο μοντέλο των χρονολογικών δεδομένων:

Heteroskedasticity Test: Breusch-Pagan-Godfrey				
Null hypothesis: Homoskedasticity				
F-statistic	0.083893	Prob. F(2,14)	0.9200	
Obs*R-squared	0.201328	Prob. Chi-Square(2)	0.9042	
Scaled explained SS	0.127501	Prob. Chi-Square(2)	0.9382	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 05/05/25 Time: 18:07				
Sample: 2010 2026				
Included observations: 17				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	11.49843	19.90700	0.577607	0.5727
GNI_AT_CONSTANT_PRICES_MRD	-0.040371	0.114523	-0.352512	0.7297
EXPORTS_OF_GOODS__AT_CONSTANT...	-0.012734	0.189264	-0.067283	0.9473
R-squared	0.011843	Mean dependent var	3.639793	
Adjusted R-squared	-0.129322	S.D. dependent var	5.127217	
S.E. of regression	5.448673	Akaike info criterion	6.387406	
Sum squared resid	415.6325	Schwarz criterion	6.534444	
Log likelihood	-51.29295	Hannan-Quinn criter.	6.402022	
F-statistic	0.083893	Durbin-Watson stat	2.122698	
Prob(F-statistic)	0.919988			

Όπως και πριν, ο έλεγχος **ΔΕΝ** εντοπίζει καμία γραμμική σχέση ανάμεσα στα τετράγωνα των καταλοίπων και τις τιμές των ανεξάρτητων μεταβλητών X_1 και X_2 .

ΓΕΝΙΚΕΥΣΗ:

$$\sigma_i^2 = a_0 + a_1 X_{1i} + a_2 X_{2i} + \dots + a_k X_{ki} + v_t$$

Θέλουμε η διακύμανση των ϵ_i να είναι **σταθερή** και ίση με σ^2 (**ομοσκεδαστικότητα**).
Επομένως, **ΔΕΝ** θέλουμε να εξαρτάται από τις τιμές των ανεξάρτητων μεταβλητών (X_i).
Παραβίαση της υπόθεσης της ομοσκεδαστικότητας συνεπάγεται την ύπαρξη της **ετεροσκεδαστικότητας**. **Συνέπειες**: ανακριβείς εκτιμητές, όχι σωστά s.e(b) → άρα, όχι σωστά συμπεράσματα στους ελέγχους υποθέσεων για την σημαντικότητα των συντελεστών, καθώς και λάθος συμπεράσματα στα διαστήματα εμπιστοσύνης.

- Τρέχουμε την παλινδρόμηση
- Αποθηκεύουμε τα κατάλοιπα (*residuals*)
- Τρέχουμε μια νέα παλινδρόμηση με εξαρτημένη μεταβλητή τα τετράγωνα των καταλοίπων και ανεξάρτητες τις μεταβλητές X .

$$\hat{u}_i^2 = a_0 + a_1 X_{1i} + a_2 X_{2i} + \dots + a_k X_{ki} + v_t$$

Ο έλεγχος του **White**:

Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic	1.754256	Prob. F(5,11)	0.2030
Obs*R-squared	7.541836	Prob. Chi-Square(5)	0.1834
Scaled explained SS	4.776245	Prob. Chi-Square(5)	0.4438

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 05/05/25 Time: 18:30
Sample: 2010 2026
Included observations: 17

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-242.4652	310.6291	-0.780562	0.4515
GNI__AT_CONSTANT_PRICES_MRD^2	-0.008030	0.008281	-0.969662	0.3531
GNI__AT_CONSTANT_PRICES_MRD*EX...	-0.042139	0.017542	-2.402096	0.0351
GNI__AT_CONSTANT_PRICES_MRD	3.616961	3.254289	1.111444	0.2901
EXPORTS_OF_GOODS__AT_CONSTANT...	0.186694	0.066513	2.806880	0.0171
EXPORTS_OF_GOODS__AT_CONSTANT...	-3.008371	3.166667	-0.950012	0.3625
R-squared	0.443637	Mean dependent var	3.639793	
Adjusted R-squared	0.190745	S.D. dependent var	5.127217	
S.E. of regression	4.612372	Akaike info criterion	6.165926	
Sum squared resid	234.0137	Schwarz criterion	6.460001	
Log likelihood	-46.41037	Hannan-Quinn criter.	6.195158	
F-statistic	1.754256	Durbin-Watson stat	2.237295	
Prob(F-statistic)	0.203025			

Είμαστε και πάλι σε θέση να **ΜΗΝ** απορρίψουμε την μηδενική υπόθεση της ομοσκεδαστικότητας και με τα 3 κριτήρια που χρησιμοποιούνται (βλ. **F-stat**, **LM**, **scaled explained SS**).

Παρατηρήστε τι ακριβώς παλινδρομεί ο έλεγχος του **White**: Ελέγχει για την ύπαρξη σχέσεων ανάμεσα στις τιμές των τετραγώνων των καταλοίπων και τις αντίστοιχες των ανεξάρτητων μεταβλητών (X_1 , X_2), αλλά και των τετραγώνων τους (X_1^2 , X_2^2), καθώς και των μεταξύ τους σχέσεων (X_1X_2). Είναι σαφές πως είναι **πιο διερευμένος** σε σχέση με τον προηγούμενο έλεγχο.

Με παρόμοιο τρόπο, **ΔΕΝ** μπορούμε να απορρίψουμε την μηδενική υπόθεση της ομοσκεδαστικότητας όταν τρέχουμε το μοντέλο:

Heteroskedasticity Test: White				
Null hypothesis: Homoskedasticity				
F-statistic	0.491604	Prob. F(5,24)	0.7792	
Obs*R-squared	2.787081	Prob. Chi-Square(5)	0.7328	
Scaled explained SS	2.188002	Prob. Chi-Square(5)	0.8226	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 05/05/25 Time: 19:07				
Sample: 1 30				
Included observations: 30				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.984277	4.100050	-0.483964	0.6328
INDUSTRY__INCLUDING_CONSTRUCTI...	-0.000406	0.003916	-0.103745	0.9182
INDUSTRY__INCLUDING_CONSTRUCTI...	0.004432	0.059365	0.074657	0.9411
INDUSTRY__INCLUDING_CONSTRUCTI...	0.054469	0.260694	0.208938	0.8363
LOG(TIME_REQUIRED_TO_START_A_B...	-0.464185	0.555248	-0.835995	0.4114
LOG(TIME_REQUIRED_TO_START_A_B...	1.792206	2.366391	0.757358	0.4562
R-squared	0.092903	Mean dependent var	0.784083	
Adjusted R-squared	-0.096076	S.D. dependent var	1.110312	
S.E. of regression	1.162426	Akaike info criterion	3.315752	
Sum squared resid	32.42962	Schwarz criterion	3.595991	
Log likelihood	-43.73628	Hannan-Quinn criter.	3.405403	
F-statistic	0.491604	Durbin-Watson stat	2.347202	
Prob(F-statistic)	0.779225			

ΓΕΝΙΚΕΥΣΗ:

Έστω ένα μοντέλο με 3 ανεξάρτητες μεταβλητές ($k=3$):

$$Y_i = \beta_1 + \beta_2 X_{1i} + \beta_3 X_{2i} + \beta_4 X_{3i} + \mu_i$$

Τρέξετε την παλινδρόμηση → Αποθηκεύστε τα κατάλοιπα → Τρέξετε μια νέα παλινδρόμηση με εξαρτημένη τα τετράγωνα των καταλοίπων και ανεξάρτητες τα x_1 , x_2 , x_3 , τα τετράγωνά τους, αλλά και τους μεταξύ τους συνδυασμούς ...

$$\mu_i^2 = \alpha_1 + \alpha_2 X_{1i} + \alpha_3 X_{2i} + \alpha_4 X_{3i} + \alpha_5 X_{1i}^2 + \alpha_6 X_{2i}^2 + \alpha_7 X_{3i}^2 + \alpha_8 X_{1i}X_{2i} + \alpha_9 X_{1i}X_{3i} + \alpha_{10} X_{2i}X_{3i} + v_i$$

Θυμηθείτε: **ΔΕΝ** θέλουμε καμία σχέση ανάμεσα στα κατάλοιπα και τις τιμές των ανεξάρτητων μεταβλητών (υπόθεση ομοσκεδαστικότητας)

$$n.R^2 \sim \chi^2_{d=9}$$

Υπολογισμός της στατιστικής τιμής ελέγχου → Σύγκριση με την κριτική τιμή → Εξαγωγή συμπερασμάτων

Πως διορθώνω την ύπαρξη ετεροσκεδαστικότητας;

Το enviews σας δίνει σαν προεπιλογή την **διόρθωση** σε περίπτωση ύπαρξης ετεροσκεδαστικότητας → μέθοδος των **Huber - White**

Στην ουσία δίνει **διορθωμένα τυπικά σφάλματα**. Έτσι, δεν ανησυχούμε για πιθανά λάθη στα συμπεράσματα στους ελέγχους στατιστικής σημαντικότητας των συντελεστών και τα διαστήματα εμπιστοσύνης.

Dependent Variable: PRIVATE_FCE__AT_CONSTANT_PRICES_MRD Method: Least Squares Date: 05/05/25 Time: 19:20 Sample: 2010 2026 Included observations: 17					Dependent Variable: PRIVATE_FCE__AT_CONSTANT_PRICES_MRD Method: Least Squares Date: 05/05/25 Time: 19:21 Sample: 2010 2026 Included observations: 17 Huber-White-Hinkley (HC1) heteroskedasticity consistent standard errors and covariance				
Variable	Coefficient	Std. Error	t-Statistic	Prob.	Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-18.51285	7.680931	-2.410236	0.0303	C	-18.51285	7.359058	-2.515655	0.0247
GNI__AT_CONSTANT_PRICES_MRD	0.759679	0.044188	17.19213	0.0000	GNI__AT_CONSTANT_PRICES_MRD	0.759679	0.044931	16.90781	0.0000
EXPORTS_OF_GOODS__AT_CONSTANT...	0.284215	0.073026	3.891996	0.0016	EXPORTS_OF_GOODS__AT_CONSTANT...	0.284215	0.082700	3.436721	0.0040
R-squared	0.967201	Mean dependent var	130.7176		R-squared	0.967201	Mean dependent var	130.7176	
Adjusted R-squared	0.962515	S.D. dependent var	10.85854		Adjusted R-squared	0.962515	S.D. dependent var	10.85854	
S.E. of regression	2.102320	Akaike info criterion	4.482745		S.E. of regression	2.102320	Akaike info criterion	4.482745	
Sum squared resid	61.87647	Schwarz criterion	4.629783		Sum squared resid	61.87647	Schwarz criterion	4.629783	
Log likelihood	-35.10333	Hannan-Quinn criter.	4.497361		Log likelihood	-35.10333	Hannan-Quinn criter.	4.497361	
F-statistic	206.4199	Durbin-Watson stat	2.226810		F-statistic	206.4199	Durbin-Watson stat	2.226810	
Prob(F-statistic)	0.000000				Prob(F-statistic)	0.000000	Wald F-statistic	256.4654	
					Prob(Wald F-statistic)	0.000000			

Παρατηρήστε την διαφορά όταν τρέχουμε το μοντέλο **χωρίς διόρθωση** για ετεροσκεδαστικότητα (αριστερά) και **με διόρθωση** (δεξιά).

Τα τυπικά σφάλματα στα αριστερά μπορεί να είναι **υποτιμημένα** (αν υπάρχει ετεροσκεδαστικότητα)→ αυτό σημαίνει πως θα οδηγούν με μεγαλύτερα ευκολία στην απόρριψη της μηδενικής υπόθεσης στον έλεγχο της στατιστικής σημαντικότητας.

Τι γίνεται όμως αν αυτή η σημαντικότητα δεν είναι πραγματική, αλλά αποτέλεσμα της υποτίμησης του πραγματικού $s.e(b)$;

Η μέθοδος των Huber-White (δεξιά) δίνει τα **διορθωμένα τυπικά σφάλματα**, που συνήθως είναι μεγαλύτερα.

.....
Διόρθωση Ετεροσκεδαστικότητας:

Heteroskedasticity Test: White				
Null hypothesis: Homoskedasticity				
F-statistic	4.277203	Prob. F(2,14)	0.0355	
Obs*R-squared	6.447738	Prob. Chi-Square(2)	0.0398	
Scaled explained SS	6.939385	Prob. Chi-Square(2)	0.0311	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 05/05/25 Time: 19:54				
Sample: 2010 2026				
Included observations: 17				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1667.081	619.9080	2.689240	0.0176
EXPORTS_OF_GOODS__AT_CONSTANT...	1.552168	0.687187	2.258728	0.0404
EXPORTS_OF_GOODS__AT_CONSTANT...	-102.2718	42.35005	-2.414916	0.0300
R-squared	0.379279	Mean dependent var	80.48341	
Adjusted R-squared	0.290604	S.D. dependent var	137.9432	
S.E. of regression	116.1835	Akaike info criterion	12.50700	
Sum squared resid	188980.6	Schwarz criterion	12.65404	
Log likelihood	-103.3095	Hannan-Quinn criter.	12.52162	
F-statistic	4.277203	Durbin-Watson stat	1.118170	
Prob(F-statistic)	0.035504			

Σ' αυτό το μοντέλο **ΑΠΟΡΡΙΠΤΟΥΜΕ** την μηδενική υπόθεση της ομοσκεδαστικότητας. Οι τιμές των p-value είναι μικρότερες από το 0,05 (5%) και για τα 3 κριτήρια που εξετάζονται. Αυτό σημαίνει πως υπάρχει **πρόβλημα ετεροσκεδαστικότητας**.

Dependent Variable: PRIVATE_FCE__AT_CONSTANT_PRICES_MRD				
Method: Least Squares				
Date: 05/05/25 Time: 19:57				
Sample: 2010 2026				
Included observations: 17				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	108.3896	9.648904	11.23336	0.0000
EXPORTS_OF_GOODS__AT_CONSTANT...	0.737470	0.309374	2.383754	0.0308
R-squared	0.274742	Mean dependent var	130.7176	
Adjusted R-squared	0.226391	S.D. dependent var	10.85854	
S.E. of regression	9.550630	Akaike info criterion	7.461222	
Sum squared resid	1368.218	Schwarz criterion	7.559247	
Log likelihood	-61.42039	Hannan-Quinn criter.	7.470966	
F-statistic	5.682283	Durbin-Watson stat	0.423051	
Prob(F-statistic)	0.030786			

Στο μοντέλο μας είμαστε σε θέση να απορρίψουμε την μηδενική υπόθεση του ελέγχου στατιστικής σημαντικότητας για τον συντελεστή β ($p\text{-value}=0,03 < 0,05$). **Αν όμως η αιτία της απόρριψης της H_0 κρύβεται στην υποτίμηση του $s.e(b)$ εξαιτίας της ύπαρξης ετεροσκεδαστικότητας ; ; ;**

Dependent Variable: PRIVATE_FCE__AT_CONSTANT_PRICES_MRD
Method: Least Squares
Date: 05/05/25 Time: 20:00
Sample: 2010 2026
Included observations: 17
Huber-White-Hinkley (HC1) heteroskedasticity consistent standard errors
and covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	108.3896	12.87421	8.419128	0.0000
EXPORTS_OF_GOODS__AT_CONSTANT...	0.737470	0.379689	1.942301	0.0711
R-squared	0.274742	Mean dependent var	130.7176	
Adjusted R-squared	0.226391	S.D. dependent var	10.85854	
S.E. of regression	9.550630	Akaike info criterion	7.461222	
Sum squared resid	1368.218	Schwarz criterion	7.559247	
Log likelihood	-61.42039	Hannan-Quinn criter.	7.470966	
F-statistic	5.682283	Durbin-Watson stat	0.423051	
Prob(F-statistic)	0.030786	Wald F-statistic	3.772532	
Prob(Wald F-statistic)	0.071118			

Πίνακας με διόρθωση ετεροσκεδαστικότητας: τώρα, το **s.e(b)** ανεβαίνει → από **0,30** πηγαίνει στο **0,37**. Αντίστοιχα, αυξάνεται και η τιμή του p-value → από **0,03** πηγαίνει στο **0,07**. **Αρχικά, απορρίπταμε την H_0 στο $\alpha=5\%$. Μετά τη διόρθωση, ΔΕΝ μπορούμε να απορρίψουμε την H_0 .** Εδώ, αλλάζει ολόκληρο το συμπέρασμα του ελέγχου.

https://en.wikipedia.org/wiki/Heteroskedasticity-consistent_standard_errors
