

Application of Equality Test of Coefficients of Variation to the Heteroskedasticity Test

Josoa Michel Tovohery¹, André Totohasina², Feno Daniel Rajaonasy³

¹Thematic Doctoral School “Science, Culture, Society and Development” of the University of Toamasina, Toamasina, Madagascar

²Department of Mathematics and Computer Science, École Normale Supérieure pour l’Enseignement Technique, University of Antsiranana, Antsiranana, Madagascar

³Department of Mathematics, Computer Science and Applications, University of Toamasina, Toamasina, Madagascar

Email: josoamicheltovoher@gmail.com, andre.totohasina@gmail.com, fenodaniel2@yahoo.fr

How to cite this paper: Tovohery, J.M., Totohasina, A. and Rajaonasy, F.D. (2020) Application of Equality Test of Coefficients of Variation to the Heteroskedasticity Test. *American Journal of Computational Mathematics*, 10, 73-89.

<https://doi.org/10.4236/ajcm.2020.101005>

Received: January 19, 2020

Accepted: March 3, 2020

Published: March 6, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The presence of heteroskedasticity in a considered regression model may bias the standard deviations of parameters obtained by the Ordinary Least Square (OLS) method. In this case, several hypothesis tests on the model under consideration may be biased, for example, CHOW’s coefficient stability test (or structural change test), Student’s t-test and Fisher’s F-test. Most of the heteroscedasticity tests in the literature are based on the comparison of variances. Despite the multiplication of equality tests of coefficients of variation (CVs) that have appeared in the literature, to our knowledge, the first and only use of the coefficient of variation in the detection of heteroskedasticity was offered by Li and Yao in 2017. Thus, this paper offers an approach to determine the existence of heteroskedasticity by a test of equality of coefficients of variation. We verify by a Monte Carlo robustness and performance test that our method seems even better than some tests in the literature. The results of this study contribute to the exploitation of the statistical measurement of CV dispersion. They help technicians economists to better verify their hypotheses before making a scientific decision when making a necessary forecast, in order to contribute effectively to the economic and sustainable development of a company or enterprise.

Keywords

Heteroskedasticity Tests, Equality Test, Coefficients of Variation, Ordinary Least Square (OLS) Method, Linear Regression, Analysis of Variance (ANOVA)

1. Introduction

Gauss-Markov’s theorem states that the least squares estimator is called BLUE,

because it is the Best linear Unbiased Estimator, in the sense that it provides the lowest variances for estimators ([1], p. 53). However, the presence of heteroskedasticity in a considered regression model may bias the standard deviations of parameters obtained by the Ordinary Least Square (OLS) method ([2], p. 31). In this case, several hypothesis tests on the model under consideration may be biased, for example, CHOW’s coefficient stability test (or structural change test) ([3], p. 25), Student’s t-test and Fisher’s F-test. Heteroscedasticity tests are already available in the literature. Examples include the Levene test, the Goldfeld-Quandt test, the White test, the Gleisjer test and the Breush test. Most of these tests are based on the comparison of variances.

Today, tests of comparison of Coefficients of Variation (CVs) have appeared in the literature. Examples include the Curto test [4], the application of the Rényi divergence proposed by Pardo (1999) [5], the test based on a numerical approach by Gokpinar (2015) [6], the Forkman test [7], McKay and Miller’s statistics [8].

To our knowledge, the first use of the coefficient of variation in the detection of heteroskedasticity was offered by Li and Yao (2017) [9]. Thus, the question is: “is it possible to find an application of these CV equality tests to detect the existence of heteroskedasticity?”

The rest of this article is organized as follows: Section 2 will discuss the position of our problem; Section 3 will present a state of the art on heteroskedasticity test; Section 4 will propose an approach to using a CV equality test when detecting heteroskedasticity; and finally, a conclusion is given at the end.

2. Position of Problem

We have a simple linear regression model

$$y_t = a_0 + a_1x_t + \epsilon_t, t = \overline{1, n} \tag{1}$$

such that the ϵ_t are the errors made when applying the model. We want to check if the variance of the errors is constant for t ranging from 1 to n . That is, we want to test if the model is homoscedastic or heteroscedastic. **Figure 1** shows an example of homoscedastic model, and **Figures 2-4** show three examples of

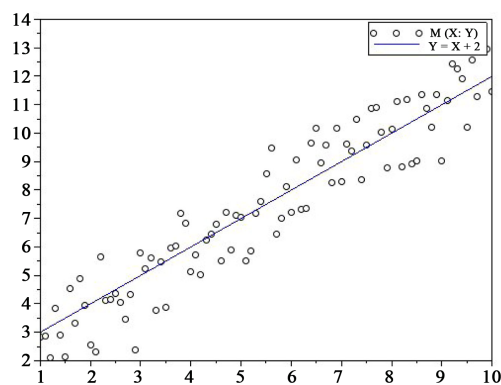


Figure 1. Homoskedastic model ($\sigma_\epsilon^2 = \text{constant}$).

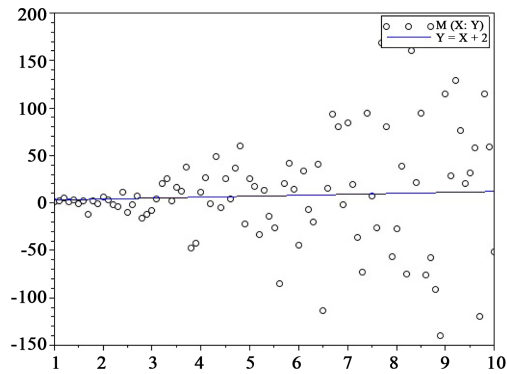


Figure 2. Heteroscedastic model (σ_{ϵ}^2 increases with the exogenous variable).

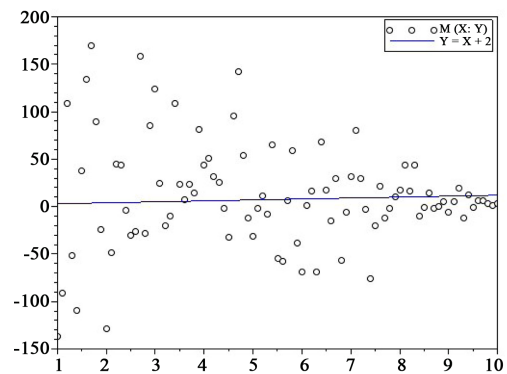


Figure 3. Heteroscedastic model (σ_{ϵ}^2 decreases with the exogenous variable).

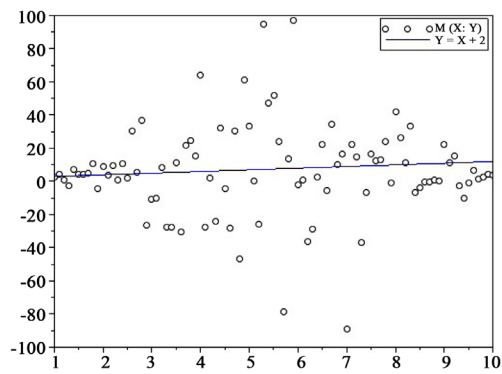


Figure 4. Heteroscedastic model (σ_{ϵ}^2 represents a concave look).

heteroscedastic model. We note that these four models all have the same regression line equation: $y = x + 2$.

3. State of the Art on the Homoskedasticity Test

We consider the general linear regression model $Y = Xa + \epsilon$. The various tests, which we will mention below, consist in testing the following hypothesis:

$$\begin{cases} \text{nul hypothesis } H_0 : \sigma_{\epsilon_t} = \sigma, t = \overline{1, n} \quad (\text{constant}); \\ \text{alternative hypothesis } H_1 : \text{there are } t_1 \text{ and } t_2, \text{ such as } \sigma_{t_1} \neq \sigma_{t_2}. \end{cases}$$

3.1. Breusch-Pagan Test

The Breusch-Pagan Test assumes that the squares of the errors ϵ_i^2 are related to the dependent variable Y . According to Leblond (2003) ([2], p. 31), the Breusch-Pagan test is done in four steps:

- 1) Recover the residues ϵ_i of the regression;
- 2) Generate the residue square (ϵ_i^2);
- 3) Regress the residue square on the variables dependent on the original regression ($\epsilon_i^2 = \hat{a}_0 + \hat{a}_1 y_i$, where \hat{a}_0 and \hat{a}_1 to be determined);
- 4) Test if the coefficients are jointly significant (Perform the F-test):

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} \tag{2}$$

where k is the number of explanatory variables x_i , n is the sample size and R^2 is the coefficient of determination of ϵ^2 and Y .

Decision-making: We accept the null hypothesis H_0 at the confidence level $(1-\alpha)\times 100\%$, if $F < F_{k;n-k-1}^\alpha$, where $F_{k;n-k-1}^\alpha$ is the critical value of F-distribution at risk α , at k and $n-k-1$ degrees of freedom.

3.2. Goldfeld-Quandt Test

The Goldfeld-Quandt test assumes that there is an explanatory variable X_i that influences the variance of errors, such as $E(\epsilon^2 | X_i) = \sigma^2 + h(X_i)$, where h is an increasing function ([10], p. 103). The test is summarized as follows:

- 1) Sort the observation values according to the increasing or decreasing values of the explanatory variable X_i suspected of being the source of heteroskedasticity.
- 2) Divide the observations into two groups:

$$Y_1 = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n_1} \end{pmatrix}, Y_2 = \begin{pmatrix} y_{n_2+1} \\ y_{n_2+2} \\ \vdots \\ y_n \end{pmatrix}$$

$$X_1 = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,k} \\ 1 & x_{2,1} & \cdots & x_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n_1,1} & \cdots & x_{n_1,k} \end{pmatrix}, X_2 = \begin{pmatrix} 1 & x_{n_2+1,1} & \cdots & x_{n_2+1,k} \\ 1 & x_{n_2+2,1} & \cdots & x_{n_2+2,k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,k} \end{pmatrix}$$

where $n_1 = n/3$ and $n_2 = 2n/3$.

- 3) Calculate the error variance estimators for each sub-sample:

$$\hat{\sigma}_1^2 = \left[(Y_1 - X_1 \hat{a})' \times (Y_1 - X_1 \hat{a}) \right] / (n_1 - k - 1) = \left(\sum_{i=1}^{n_1} e_i^2 \right) / (n_1 - k - 1) \tag{3}$$

$$\begin{aligned} \hat{\sigma}_2^2 &= \left[(Y_2 - X_2 \hat{a})' \times (Y_2 - X_2 \hat{a}) \right] / (n - n_2 - k - 1) \\ &= \left(\sum_{i=n_2+1}^n e_i^2 \right) / (n - n_2 - k - 1) \end{aligned} \tag{4}$$

where \hat{a} is the estimator of the parameter a by the least squares method,

$e_i = y_i - (\hat{a}_0 + \hat{a}_1x_{i1} + \dots + \hat{a}_kx_{ik})$ and k is the number of explanatory variables of the model.

4) Calculate the Goldfeld-Quandt statistic:

$$GQ = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \tag{5}$$

The GQ statistic follows the F -distribution at $n_1 - k - 1$ and $n - n_2 - k - 1$ degrees of freedom, noted as $F_{n_1 - k - 1; n - n_2 - k - 1}$.

Decision-making: The null hypothesis H_0 is rejected at confidence level $(1 - \alpha) \times 100\%$, if $GQ > F_{n_1 - k - 1; n - n_2 - k - 1; \alpha}$.

3.3. Gleisjer's Test

The Gleisjer test can detect both heteroskedasticity and the form that this heteroskedasticity takes ([1], p. 150). The Gleisjer test assumes that there is a relationship between the error ϵ of the model and the variable X_i assumed to be the cause of heteroskedasticity. The steps of the test are summarized as follows:

Step 1: Determination of the residues generated by the suspected variable X_i .

1) Regress Y to X . This gives the simple regression model

$$Y_k = aX_{ki} + b + \epsilon_k, k = \overline{1, n}.$$

2) Calculate the estimators of a and b using the Ordinary Least Squares method: \hat{a} and \hat{b} .

3) Estimate the model's residues ϵ_k by its estimators:

$$e_k = Y_k - (\hat{a}X_{ki} + \hat{b}), k = \overline{1, n}.$$

Thus, the vector of residues e_k is known.

Step 2: Proposal of possible forms of existing heteroskedasticity.

Gleisjer suggests testing different forms of possible relationships between $|e|$ and X_i , for example:

1) Type 1:

$$|e_k| = a_0 + a_1X_{ki} + v_k, k = \overline{1, n}, \tag{6}$$

where v_k is the residue of this model. This relationship generates the type of heteroskedasticity $\hat{\sigma}_{e_k}^2 = c^2 X_{ki}^2$, where c is a non-zero real constant. Thus, the variance of errors is a function of the squares of the suspected explanatory variable X_i .

2) Type 2:

$$|e_k| = a_0 + a_1\sqrt{X_{ki}} + v_i, k = \overline{1, n}. \tag{7}$$

This relationship generates the type of heteroskedasticity $\hat{\sigma}_{e_k}^2 = c^2 X_{ki}$. In this case, the variance of the errors is proportional to the values of the suspected explanatory variable X_i .

3) Type 3:

$$|e_k| = a_0 + a_1\frac{1}{X_{ki}} + v_i, k = \overline{1, n}. \tag{8}$$

This relationship leads to heteroskedasticity of type $\hat{\sigma}_{e_k}^2 = \frac{c^2}{X_{ki}^2}$.

Step 3: Detection of heteroskedasticity

Significance test of the regression coefficient a_1 :

$$t^* = \frac{|\hat{a}_1|}{\hat{\sigma}_{\hat{a}_1}}, \tag{9}$$

with

$$\hat{\sigma}_{\hat{a}_1} = \sqrt{\frac{\hat{\sigma}_v^2}{\sum_{k=1}^n (h(X_{ki}) - \bar{h}(X_i))^2}}$$

and

$$\hat{\sigma}_v^2 = \frac{1}{n-2} \sum_{k=1}^n [|e_k| - (\hat{a}_0 + \hat{a}_1 h(X_{ki}))]^2,$$

where

$$h(x) = \begin{cases} x, & \text{for a type 1 relation ship;} \\ \sqrt{x}, & \text{for a type 2 relation ship;} \\ \frac{1}{x}, & \text{for a type 3 relation ship.} \end{cases}$$

t^* follows the t-distribution at $n - 2$ degrees of freedom.

Decision-making: The null hypothesis H_0 is rejected at confidence level $(1 - \alpha) \times 100\%$, if there is a t^* , such that $t^* > t_{n-2, \alpha}$.

If the existence of heteroskedasticity is validated, then the relationship with the highest t^* represents the form of existing heteroskedasticity.

3.4. White’s Test

White’s test consists in testing the existence of a relationship between the square of the residue and one or more explanatory variables or its squares. The test procedures can be summarized as follows:

Step 1: Determination of model’s residues.

1) When the parameters of the model $Y = Xa + \epsilon$ are estimated, then we have the estimation of the residues: $e = Y - X\hat{a}$.

2) **Step 2:** Regression of e^2 to x_1, x_1^2, \dots, x_k and x_k^2 and validation.

3) We consider the model:

$$e_i^2 = a_1 x_{i1} + b_1 x_{i1}^2 + a_2 x_{i2} + b_2 x_{i2}^2 + \dots + a_k x_{ik} + b_2 x_{ik}^2 + a_0 + v_i, i = \overline{1, n}, \tag{10}$$

what can be written in matrix form: $E = Wu + v$, where

$$E = \begin{pmatrix} e_1^2 \\ e_2^2 \\ \vdots \\ e_n^2 \end{pmatrix}, u = \begin{pmatrix} a_0 \\ a_1 \\ b_1 \\ \vdots \\ a_n \\ b_n \end{pmatrix}, v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \text{ and } W = \begin{pmatrix} 1 & x_{11} & x_{12}^2 & \dots & x_{1k} & x_{1k}^2 \\ 1 & x_{21} & x_{22}^2 & \dots & x_{2k} & x_{2k}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{n1} & x_{n1}^2 & \dots & x_{nk} & x_{nk}^2 \end{pmatrix}$$

4) The estimator of u is: $\hat{u} = (W' \cdot W)^{-1} \cdot W' \cdot E$

5) Calculate the variance of the errors: $\hat{\sigma}_\epsilon^2 = \frac{\sum_{i=1}^n \hat{v}_i^2}{n - k - 1}$, with $\hat{v} = E - W\hat{u}$.

6) Calculate the variance-covariance matrix of parameters a_i and b_i : $\hat{\Omega}_u = \hat{\sigma}_\epsilon^2 (W' \cdot W)^{-1}$.

In this case, the variance of i -th element of the vector u is: $\hat{\sigma}_{\hat{u}_i} = i$ -th element of the diagonal of $\hat{\Omega}_u$.

7) Significance test of parameters $a_1, b_1, \dots, a_k, b_k$: We calculate: $t_{a_i}^* = \frac{|\hat{a}_i|}{\hat{\sigma}_{\hat{a}_i}}$ and

$$t_{b_i}^* = \frac{|\hat{b}_i|}{\hat{\sigma}_{\hat{b}_i}}, \quad i = \overline{1, k}.$$

The statistics $t_{a_i}^*$ and $t_{b_i}^*$ follow the t-distribution at $n - k - 1$ degrees of freedom.

Decision-making: The null hypothesis H_0 is rejected at the confidence level $(1 - \alpha) \times 100\%$, if there is a $t_{u_i}^*$, such that $t_{u_i}^* > t_{n-k-1; \alpha}$. That means, the null hypothesis H_0 is rejected if there is a parameter u_i significantly different from 0.

3.5. ANOVA Methods

In order to determine the existence of heteroskedasticity, researchers proposed the method of analysis of variances, commonly said ANOVA. According to the application example presented in ([1], p. 147-148), the application of ANOVA consists in dividing the observations into several classes of values. Following the example of this same example by R. Bourbonnais, we propose the following steps:

1) Order the observations according to the increasing values of the explanatory variable X_i suspected to be the source of heteroskedasticity.

2) Group the value of the variable X_i into z classes of values. To determine z , one of the following expressions can be used in ([11], p. 33):

a) $z = \text{Int}(\sqrt{n})$, where n is the total number of observations, and $\text{Int}(\cdot)$ is the integer part function;

b) Sturge's formula: $z = \text{Int}(1 + 3.3 \log_{10}(n))$;

c) Yule's formula: $z = \text{Int}(2.5 \sqrt[4]{n})$.

3) Group the values of the variable to be explained Y according to their corresponding classes (y_i in the class corresponding to x_i). Thus, we obtain z samples of Y .

4) Apply the ANOVA test to the z samples of Y , then draw a conclusion.

In the following subsections, we will present some ANOVA tests that can be done in step 4.

3.5.1. Bartlett's Test

Bartlett's statistic¹ is defined as follows:

$$B = \frac{Q}{L} \tag{11}$$

¹Maurice Stevenson Bartlett (June 18, 1910-January 8, 2002).

where

$$Q = (n - z) \ln \left(\sum_{i=1}^z \frac{n_i - 1}{n - z} s_i^2 \right) - \sum_{i=1}^z (n_i - 1) \ln(s_i^2),$$

$$L = 1 + \frac{1}{3(z-1)} \left(\sum_{i=1}^z \frac{1}{n_i - 1} - \frac{1}{n - z} \right),$$

$n = \sum_{i=1}^z n_i$ and n_i is the number of observations belonging to the i -th class, $i = \overline{1, z}$ ([12], p. 273).

Remark: Bartlett's statistic B follows the chi-square distribution with $z - 1$ degrees of freedom, noted as χ_{z-1}^2 , if the residues ϵ_i are independent and follow the standard normal distribution $\mathcal{N}(0,1)$.

Decision-making: The homoskedasticity hypothesis H_0 is rejected at confidence level $(1 - \alpha) \times 100\%$, if $Q \geq \chi_{z-1; 1-\alpha}^2$.

3.5.2. Levene's Test

The Howard Levene's statistic proposed in 1960 ([13], p. 4) is defined as follows:

$$F = \frac{n - z}{z - 1} \times \frac{\sum_{i=1}^z (\bar{d}_i - \bar{d}_{..})^2}{\sum_{i=1}^z \sum_{j=1}^{n_i} n_i (d_{ij} - \bar{d}_i)^2} \tag{12}$$

where,

- z is the number of groups or value categories obtained,
- n_i is the number of observations belonging to the i -th class, and

$$n = \sum_{i=1}^z n_i,$$

- $d_{ij} = |y_{ij} - \bar{y}_i|$,

- $d_i = \frac{1}{n_i} \sum_{j=1}^{n_i} d_{ij}$ (average of d_{ij} in the i class),

- $d_{..} = \frac{1}{n} \sum_{i=1}^z \sum_{j=1}^{n_i} d_{ij}$ (average of all d_{ij}).

Remark: Levene's F statistic follows the F-distribution with $z - 1$ and $n - z$ degrees of freedom, noted $F_{z-1; n-z}$. Bartlett's test is not robust if the normality assumption of ϵ_i is not verified. However, the Levene test is stable even in the absence of this hypothesis.

Decision making: The null hypothesis H_0 is rejected at the confidence level $(1 - \alpha) \times 100\%$ if $F > F_{z-1; n-z; \alpha}$.

3.5.3. Brown-Forsythe's Test

The Brown-Forsythe test is an improvement on the Levene test. To get the Brown-Forsythe statistic, just change $d_{ij} = |y_{ij} - \bar{y}_i|$ to $d_{ij} = |y_{ij} - me_i|$, where me_i is the median of the i -th group of values. Brown-Forsythe's statistic is more robust than Levene's.

3.5.4. Hartley's Test

We define the Hartley's statistic ([14], p. 14) by:

$$H = \frac{s_{\max}^2}{s_{\min}^2} \tag{13}$$

where $s_{\max}^2 = \max\{s_1^2, \dots, s_z^2\}$, $s_{\min}^2 = \min\{s_1^2, \dots, s_z^2\}$ and $s_i^2 =$ variance of the Y values of the i -th group, such as $i = 1, 2, \dots, z$.

Remark: The Hartley test cannot be used if the group sizes n_i are not equal. The critical values of the H statistic are tabulated in the Hartley table.

Decision making: We reject null hypothesis H_0 at the confidence level $(1 - \alpha) \times 100\%$ if $H > H_{critical}(\alpha)$

3.5.5. Cochran's Test

The Cochran's statistic is defined as follows:

$$C = \frac{s_{\max}^2}{\sum_{i=1}^z s_i^2} \tag{14}$$

Remarks: The Cochran's test cannot be used if the group sizes n_i are not equal. The critical values of the C statistic are tabulated in the Cochran's table.

Decision making: We reject the null hypothesis H_0 at the confidence level $(1 - \alpha)\%$ if $C > C_{critical}(\alpha)$.

3.6. Zhaoyuan Li and Jianfeng Yao Test

Zhaoyuan Li and Jianfeng Yao [9] proposed two measures to detect heteroskedasticity in a multivariate linear model.

1) Test based on the likelihood ratio:

$$T_1 = \ln \left(\frac{\frac{1}{n} \sum_{i=1}^n \hat{\epsilon}_i^2}{\left(\prod_{i=1}^n \hat{\epsilon}_i^2 \right)^{1/n}} \right) \tag{15}$$

where $\hat{\epsilon} = Y - X\hat{a}$ and $\hat{a} = (X'X)^{-1} X'Y$ ([9], p. 9).

$Z_1 = \frac{\sqrt{n}(T_1 - [\ln(2) + \gamma])}{\frac{\pi^2}{2} - 2}$ follows the standard normal distribution $\mathcal{N}(0,1)$,

and $\gamma \sim 0.5772$ is the Euler's constant ([9], p. 10).

Decision making: the H_0 assumption is rejected at the confidence level $(1 - \alpha) \times 100\%$, if $Z_1 > z_{\alpha/2}$, where $z_{\alpha/2}$ is the quantile of $\mathcal{N}(0,1)$ at the risk threshold α . For $\alpha = 0.05$, we have $z_{\alpha/2} = 1.96$.

2) Coefficient of variation test:

$$T_2 = \frac{\frac{1}{n} \sum_{i=1}^n (\hat{\epsilon}_i^2 - \bar{\epsilon})^2}{\bar{\epsilon}^2} \tag{16}$$

where $\bar{\epsilon} = \frac{1}{n} \sum_{i=1}^n \hat{\epsilon}_i^2$.

$Z_2 = \frac{\sqrt{n}(T_2 - 2)}{24}$ follows the standard normal distribution $\mathcal{N}(0,1)$ ([9], p. 11).

Decision making: the H_0 assumption is rejected at the confidence level $(1-\alpha)\times 100\%$, if $Z_2 > z_{\alpha/2}$.

This last test shows a trend in the use of coefficient of variation in the detection of heteroskedasticity.

4. Application of the Equality Test of Coefficients of Variation to the Heteroskedasticity Test

4.1. Our Approach

In this section, we will show that the test of equality of coefficients of variation allows us to detect the existence of heteroskedasticity. The steps of our approach can be summarized as follows:

- 1) Estimate the parameter a of the regression model of Y to X , noted as \hat{a} .
- 2) Estimate the model's residues: $\hat{\epsilon} = Y - X\hat{a}$.
- 3) Calculate the square of residues: $\hat{\epsilon}^2$.
- 4) As the Goldfeld-Quandt method, divide the residue squares into two groups:

$$\hat{\epsilon}_1^2 = \begin{pmatrix} e_1^2 \\ e_2^2 \\ \vdots \\ e_{n_1}^2 \end{pmatrix} \text{ et } \hat{\epsilon}_2^2 = \begin{pmatrix} e_{n_2+1}^2 \\ e_{n_2+2}^2 \\ \vdots \\ e_n^2 \end{pmatrix}$$

where $n_1 = n/3$ and $n_2 = 2n/3$.

- 5) Calculate the Johannes Forkman's statistic ([7], p. 10):

$$F^* = \frac{c_1^2 / [1 + c_1^2(n_1 - 1)/n_1]}{c_2^2 / [1 + c_2^2(n_2 - 1)/n_2]} \tag{17}$$

where $c_i = s_i/m_i$ for $i = \overline{1,2}$,

$$m_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} e_i^2, \quad m_2 = \frac{1}{n - n_2 + 1} \sum_{i=n_2}^n e_i^2,$$

$$s_1 = \sqrt{\frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (e_i^2 - m_1)^2} \quad \text{and} \quad s_2 = \sqrt{\frac{1}{n - n_2 + 1} \sum_{i=n_2}^n (e_i^2 - m_2)^2}.$$

Decision making: if $F^* < F_{(n_1-1, n_2-1, \alpha)}$, then we accept H_0 at the confidence level $(1-\alpha)\times 100\%$. $F_{(n_1-1, n_2-1, \alpha)}$ is the quantile α of F-distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom.

We chose Forkman's statistic because F^* is stable for all $n_i \geq 3$, where $i = \overline{1,2}$ ([7], p. 11).

4.2. Monte Carlo Simulation

Now, we will test the robustness of these measures proposed in the literature and the one in which we have proposed.

4.2.1. Methodology

Like the Gleisjer method, our simulation consists of generating two variables X

and Y of size $n = \{15; 30; 40; 50; 80; 100\}$, such as $Y = b + aX + \epsilon$ and $\epsilon = a_0 + a_1 \cdot h(X)$ (see the Section 3.3). Thus, we consider 3 forms of heteroskedasticity: 1) $h(X) = X$, 2) $h(X) = \sqrt{X}$ and 3) $h(X) = 1/X$ ([1], p. 151).

Moreover, in order to enrich the forms of heteroskedasticity studied, we also propose to take the other three forms considered by Li and Yao: 4) $h(X) = g_i \exp(cX_i)$, 5) $h(X) = g_i (1 + c \sin(10X_i))^2$ and 6) $h(X) = g_i (1 + cX_i)^2$, where g_i is a random variable following the standard normal distribution $\mathcal{N}(0,1)$ ([9], p. 15).

In this simulation, we consider only the simple regression model. We repeat $m = 100$ times this test, and we count the number k of times the test rejects the H_0 hypothesis at the 95% confidence level. Then, the probability $p = k/m$ is calculated.

As p is a random variable, then we repeat these procedures several times (1000 times), then we calculate $\bar{p} = (\sum_{i=1}^{1000} p_i) / 1000$. We really put ourselves in the case where the error is significantly not negligible (value of a_1 sufficiently different from 0).

So, if $\bar{p} > 0.05$, then the test is considered robust. In addition, the measure with the highest \bar{p} is the measure considered most sensitive to the type of error i considered ($i = \overline{1,6}$).

As we want to test the robustness of the test, then it would be better to check whether the test in question detects small variations or not. During the simulations we did, we took $a = 3$, $b = 2$, $a_0 = 2$ and $a_1 = c = 1$. We took $a_1 = 1$, because it is already different from 0, but judged subjectively low value.

In **Table 1**, the probabilities $p_1, p_2, p_3, p_4, p_5, p_6, p_7$ and p_8 correspond respectively to the rejection probabilities of the null hypothesis H_0 of the Breush, Goldfeld-Quandt, Gleisjer, White, Bartlette, Levene, Li and Yao tests, and our proposal.

4.2.2. Simulation Results

From **Table 1**, we obtain the classifications in **Tables 2-6**.

4.3. Discussion

First of all, from these simulations, it is indisputable that the Levene test is the most robust and sensitive of all the tests considered in this study.

However, these results show that, among the 06 forms of heteroskedasticity proposed, our proposal can detect 04 for $n < 50$, and 05 for $n \geq 50$.

In general, our proposal fails to detect the only form of heteroskedasticity $h(X) = g(1 + c \sin(10X))^2$ (whether for $n < 50$ or $n \geq 50$).

Furthermore, it is the second best test to detect the heteroskedasticity of type $h(X) = 1/X$ for $n \geq 50$.

In addition, our proposal seems better than the Li and Yao test, which is, to our knowledge, the first tendency to use the coefficient of variation to detect heteroskedasticity.

Table 1. Results of monte carlo simulations.

Formes	$n = 15$	$n = 30$	$n = 40$	$n = 50$	$n = 80$	$n = 100$
$h(X) = X$	$p_1 = 0.5947$	$p_1 = 0.79422$	$p_1 = 0.86975$	$p_1 = 0.91714$	$p_1 = 0.97982$	$p_1 = 0.99321$
	$p_2 = 0.36198$	$p_2 = 0.99374$	$p_2 = 0.99953$	$p_2 = 0.99996$	$p_2 = 1$	$p_2 = 1$
	$p_3 = 0.75472$	$p_3 = 0.96118$	$p_3 = 0.97529$	$p_3 = 0.98452$	$p_3 = 0.98897$	$p_3 = 0.99228$
	$p_4 = 0.17256$	$p_4 = 0.48818$	$p_4 = 0.60891$	$p_4 = 0.70495$	$p_4 = 0.87512$	$p_4 = 0.93361$
	$p_5 = 0.17245$	$p_5 = 0.54363$	$p_5 = 0.88649$	$p_5 = 0.99279$	$p_5 = 1$	$p_5 = 1$
	$p_6 = 0.85861$	$p_6 = 0.99987$	$p_6 = 1$	$p_6 = 1$	$p_6 = 1$	$p_6 = 1$
	$p_7 = 0$	$p_7 = 0.00007$	$p_7 = 0$	$p_7 = 0.00002$	$p_7 = 0.00001$	$p_7 = 0$
	$p_8 = 0$	$p_8 = 0.03175$	$p_8 = 0.05503$	$p_8 = 0.09824$	$p_8 = 0.30432$	$p_8 = 0.54455$
$h(X) = \sqrt{X}$	$p_1 = 0.36751$	$p_1 = 0.79705$	$p_1 = 0.91274$	$p_1 = 0.9628$	$p_1 = 0.99788$	$p_1 = 0.99972$
	$p_2 = 0.04419$	$p_2 = 0.17683$	$p_2 = 0.61035$	$p_2 = 0.9125$	$p_2 = 0.99891$	$p_2 = 0.99992$
	$p_3 = 0.40739$	$p_3 = 0.84986$	$p_3 = 0.94284$	$p_3 = 0.97647$	$p_3 = 0.998$	$p_3 = 0.99821$
	$p_4 = 0.10525$	$p_4 = 0.2614$	$p_4 = 0.41659$	$p_4 = 0.57791$	$p_4 = 0.88057$	$p_4 = 0.9524$
	$p_5 = 0.01851$	$p_5 = 0.01275$	$p_5 = 0.01478$	$p_5 = 0.01782$	$p_5 = 0.00947$	$p_5 = 0.01216$
	$p_6 = 0.98207$	$p_6 = 1$	$p_6 = 1$	$p_6 = 1$	$p_6 = 1$	$p_6 = 1$
	$p_7 = 0$	$p_7 = 0$	$p_7 = 0$	$p_7 = 0$	$p_7 = 0$	$p_7 = 0$
	$p_8 = 0$	$p_8 = 0.21218$	$p_8 = 0.41056$	$p_8 = 0.63869$	$p_8 = 0.91821$	$p_8 = 0.97027$
$h(X) = 1/X$	$p_1 = 0.09467$	$p_1 = 0.09515$	$p_1 = 0.09298$	$p_1 = 0.09494$	$p_1 = 0.09031$	$p_1 = 0.08729$
	$p_2 = 0.01856$	$p_2 = 0.00989$	$p_2 = 0.00966$	$p_2 = 0.00874$	$p_2 = 0.00812$	$p_2 = 0.00767$
	$p_3 = 0.10965$	$p_3 = 0.14892$	$p_3 = 0.16854$	$p_3 = 0.18732$	$p_3 = 0.21713$	$p_3 = 0.22787$
	$p_4 = 0.08128$	$p_4 = 0.09837$	$p_4 = 0.10196$	$p_4 = 0.10286$	$p_4 = 0.10383$	$p_4 = 0.10399$
	$p_5 = 0.00045$	$p_5 = 0$	$p_5 = 0$	$p_5 = 0$	$p_5 = 0$	$p_5 = 0$
	$p_6 = 0.99999$	$p_6 = 1$	$p_6 = 1$	$p_6 = 1$	$p_6 = 1$	$p_6 = 1$
	$p_7 = 0$	$p_7 = 0$	$p_7 = 0$	$p_7 = 0$	$p_7 = 0$	$p_7 = 0$
	$p_8 = 0$	$p_8 = 0.06337$	$p_8 = 0.13264$	$p_8 = 0.25046$	$p_8 = 0.6156$	$p_8 = 0.78673$
$h(X) = g \exp(cX)$	$p_1 = 0.83148$	$p_1 = 0.91736$	$p_1 = 0.93403$	$p_1 = 0.94213$	$p_1 = 0.95797$	$p_1 = 0.96288$
	$p_2 = 0.98694$	$p_2 = 1$	$p_2 = 1$	$p_2 = 1$	$p_2 = 1$	$p_2 = 1$
	$p_3 = 0.64155$	$p_3 = 0.73515$	$p_3 = 0.76323$	$p_3 = 0.75063$	$p_3 = 0.81769$	$p_3 = 0.73387$
	$p_4 = 0.53062$	$p_4 = 0.68474$	$p_4 = 0.74005$	$p_4 = 0.7433$	$p_4 = 0.81738$	$p_4 = 0.73387$
	$p_5 = 1$	$p_5 = 1$	$p_5 = 1$	$p_5 = 1$	$p_5 = 1$	$p_5 = 1$

Continued

	$p_6 = 1$	$p_6 = 1$	$p_6 = 1$	$p_6 = 1$	$p_6 = 1$	$p_6 = 1$
	$p_7 = 0$	$p_7 = 0.85231$	$p_7 = 0.98921$	$p_7 = 0.99939$	$p_7 = 1$	$p_7 = 1$
	$p_8 = 0$	$p_8 = 0.01646$	$p_8 = 0.02289$	$p_8 = 0.05477$	$p_8 = 0.98939$	$p_8 = 0.99824$
$h(X) = g(1 + c \sin(10X))^2$	$p_1 = 0.04149$	$p_1 = 0.02013$	$p_1 = 0.01793$	$p_1 = 0.01863$	$p_1 = 0.02393$	$p_1 = 0.02117$
	$p_2 = 0.12345$	$p_2 = 0.24139$	$p_2 = 0.26078$	$p_2 = 0.27889$	$p_2 = 0.27596$	$p_2 = 0.29125$
	$p_3 = 0.06083$	$p_3 = 0.03334$	$p_3 = 0.02658$	$p_3 = 0.02517$	$p_3 = 0.02429$	$p_3 = 0.02068$
	$p_4 = 0.00891$	$p_4 = 0.01691$	$p_4 = 0.01527$	$p_4 = 0.01843$	$p_4 = 0.03743$	$p_4 = 0.02913$
	$p_5 = 0$	$p_5 = 0$	$p_5 = 0$	$p_5 = 0$	$p_5 = 0$	$p_5 = 0$
	$p_6 = 1$	$p_6 = 1$	$p_6 = 1$	$p_6 = 1$	$p_6 = 1$	$p_6 = 1$
	$p_7 = 0$	$p_7 = 0.02105$	$p_7 = 0.04922$	$p_7 = 0.08594$	$p_7 = 0.17746$	$p_7 = 0.28$
	$p_8 = 0$	$p_8 = 0.00058$	$p_8 = 0.00057$	$p_8 = 0.00023$	$p_8 = 0.00002$	$p_8 = 0.00001$
$h(X) = g(1 + cX)^2$	$p_1 = 0.28919$	$p_1 = 0.39289$	$p_1 = 0.41932$	$p_1 = 0.44054$	$p_1 = 0.47065$	$p_1 = 0.48058$
	$p_2 = 0.92111$	$p_2 = 0.99998$	$p_2 = 1$	$p_2 = 1$	$p_2 = 1$	$p_2 = 1$
	$p_3 = 0.60642$	$p_3 = 0.8823$	$p_3 = 0.83984$	$p_3 = 0.91048$	$p_3 = 0.85081$	$p_3 = 0.94033$
	$p_4 = 0.11691$	$p_4 = 0.23252$	$p_4 = 0.30714$	$p_4 = 0.40276$	$p_4 = 0.5692$	$p_4 = 0.72512$
	$p_5 = 0.93684$	$p_5 = 0.99999$	$p_5 = 1$	$p_5 = 1$	$p_5 = 1$	$p_5 = 1$
	$p_6 = 0.94399$	$p_6 = 0.99965$	$p_6 = 1$	$p_6 = 0.99999$	$p_6 = 1$	$p_6 = 1$
	$p_7 = 0$	$p_7 = 0.01478$	$p_7 = 0.04695$	$p_7 = 0.08402$	$p_7 = 0.21858$	$p_7 = 0.31826$
	$p_8 = 0$	$p_8 = 0.02982$	$p_8 = 0.07841$	$p_8 = 0.13746$	$p_8 = 0.16981$	$p_8 = 0.14072$

$p_1 =$ Breush, $p_2 =$ Goldfeld-Quandt, $p_3 =$ Gleisjer,
 $p_4 =$ White, $p_5 =$ Bartlette, $p_6 =$ Levene, $p_7 =$ Li et Yao et, $p_8 =$ notre proposition.

Table 2. Classification of tests in ascending order according to their wrong acceptance numbers of H_0 .

Test	Number of times the test incorrectly accepts H_0
lightaqua	
Levene	00
Goldfeld - Quandt	01
Gleisjer	01
Breush	02
White	02
Our proposal	03
Bartlette	06
Li et Yao	08

Table 3. Classification in ascending order of tests according to their sensitivities to the 03 types of heteroskedasticity proposed by Gleisjeir for $n < 50$.

Forms	for $n < 50$		
	$h(X) = X$	$h(X) = \sqrt{X}$	$h(X) = 1/X$
classification in ascending order of test sensitivity (*: the test accepts H_0)	1) Levene	1) Levene	1) Levene
	2) Goldfeld-Quandt	2) Gleisjer	2) Gleisjer
	3) Gleisjer	3) Breush	3) Breush
	4) Breush	4) White	4) White
	5) Bartlette	5) Goldfeld-Quandt	5) Our proposal
	6) White	6) Our proposal	6) Goldfeld-Quandt*
	7) Our proposal	7) Bartlette*	7) Bartlette*
	8) Li and Yao*	8) Li and Yao*	8) Li and Yao*

Table 4. Classification in ascending order of tests according to their sensitivities to the 03 types of heteroskedasticity proposed by Gleisjeir for $n \geq 50$.

Forms	For $n \geq 50$		
	$h(X) = X$	$h(X) = \sqrt{X}$	$h(X) = 1/X$
classification in ascending order of test sensitivity (*: the test accepts H_0)	1) Levene	1) Levene	1) Levene
	2) Goldfeld-Quandt	2) Gleisjer	2) Our proposal
	3) Bartlette	3) Breush	3) Gleisjer
	4) Gleisjer	4) Goldfeld-Quandt	4) White
	5) Breush	5) Our proposal	5) Breush
	6) White	6) White	6) Goldfeld-Quandt*
	7) Our proposal	7) Bartlette*	7) Bartlette* and
	8) Li and Yao*	8) Li and Yao*	8) Li & Yao*

Table 5. Classification in ascending order of tests according to their sensitivities to the 03 types of heteroskedasticity considered by Li and Yao for $n < 50$.

Forms	For $n < 50$		
	$h(X) = g \cdot \exp(cX)$	$h(X) = g \cdot (1 + c \sin(10X))^2$	$h(X) = g \cdot (1 + cX)^2$
Classification in ascending order of test sensitivity (*: the test accepts H_0)	1) Levene and Bartlette	1) Levene	1) Levene
	2) Goldfeld-Quandt	2) Goldfeld-Quandt	2) Bartlette
	3) Breush	3) Gleisjer	3) Goldfeld-Quandt
	4) Gleisjer	4) Breush*	4) Gleisjer
	5) White	5) Li and Yao*	5) Breush
	6) Li and Yao	6) White*	6) White
	7) Our proposal*	7) Our proposal*	7) Our proposal
		8) Bartlette*	8) Li and Yao*

Table 6. Classification in ascending order of tests according to their sensitivities to the 03 types of heteroskedasticity considered by Li and Yao for $n \geq 50$.

Forms	For $n \geq 50$		
	$h(X) = g \cdot \exp(cX)$	$h(X) = g \cdot (1 + c \sin(10X))^2$	$h(X) = g \cdot (1 + cX)^2$
Classification in ascending order of test sensitivity (*: the test accepts H_0)	1) Levene, Bartlette and Goldfeld-Quandt	1) Levene	1) Bartlette, Goldfeld-Quandt and Levene
	2) Li and Yao	2) Goldfeld-Quandt	2) Gleisjer
	3) Breush	3) Li and Yao	3) White
	4) Gleisjer	4) White*	4) Breush
	5) White	5) Gleisjer*	5) Li and Yao
	6) Our proposal	6) Breush*	6) Our proposal
		7) Our proposal*	
		8) Bartlette*	

Finally, these results contribute to the justification of the weakness of Bartlette’s test. Indeed, we see from these results that this test is less robust than our proposal.

5. Conclusions

In this paper, we proposed a technique to detect the existence of heteroskedasticity by an equality test of the coefficients of variation. Thus, to illustrate our state of the art, we first recalled some tests to detect the existence of heteroskedasticity existing in the literature, such as the Breusch-Pagan test, the Goldfeld-Quandt test, the Gleisjer test, the White test and some heteroskedasticity tests based on an analysis of variance (ANOVA): Bartlett’s test, Levene’s test, Brown-Forsythe’s test, Hartley’s test and Cochran’s test.

Next, we also presented the heteroskedasticity test of Zhaoyuan Li and Jianfeng Yao. To the best of our knowledge, the Zhaoyuan Li and Jianfeng Yao test was the first tendency to use coefficients of variation to determine the existence of heteroskedasticity.

Among the equality tests of coefficients of variation available in the literature, we have considered Forkman’s test to illustrate our approach, as it is a robust and stable test for a sample with size $n \geq 3$. The results of our performance tests have shown that our approach can detect 5 types of heteroskedasticity among the 6 types considered in this paper.

At the end of this analysis, we affirm that the equality test of coefficients of variation allows us to detect the existence of possible heteroskedasticity in a simple regression model. Thus, our study contributes to the reapplication of several equality tests of coefficients of variation that have already appeared in the literature.

Acknowledgements

We thank the Editor and the referee for their comments and assistance.

Conflicts of Interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] Bourbonnais, R. (2015) Économétrie-Cours et exercices corrigés. Dunod 9^e édition. <https://docplayer.fr/66598708-Econometrie-cours-et-exercices-corriges-regis-bourbonnais-9-e-edition.html>
- [2] Leblond, S. (2003) Guide d'économétrie appliquée. Université de Montréal, Montréal. <http://www2.cirano.qc.ca/~mccauslw/ECN3949/GuideEconometrie.pdf>
- [3] Hamisultane, H. (2002) Économétrie. Licence. France. <https://halshs.archives-ouvertes.fr/cel-01261163>
- [4] Curto, J.D. and Pinto, J.C. (2009) The Coefficient of Variation Asymptotic Distribution in the Case of Non-IID Random Variables. *Journal of Applied Statistics*, **36**, 21-32. <https://doi.org/10.1080/02664760802382491>
- [5] Pardo, M.C. and Pardo, J.A. (2000) Use of $R\{e\}nyi$ Divergence to Test for the Equality of the Coefficients of Variation. *Journal of Computational and Applied Mathematics*, **116**, 93-104. [https://doi.org/10.1016/S0377-0427\(99\)00312-X](https://doi.org/10.1016/S0377-0427(99)00312-X)
- [6] Gokpinar, G. and Esra, G. (2015) A Computational Approach for Testing Equality of Coefficients of Variation in k Normal Populations. *Hacetatepe Journal of Mathematics and Statistics*, **44**, 1197-1213. <https://doi.org/10.15672/HJMS.2014317482>
- [7] Krishnamoorthy, K. and Meesook, L. (2013) Improved Tests for the Equality of Normal Coefficients of Variation. In: *Computational Statistics*, Springer, New York. <https://doi.org/10.1007/s00180-013-0445-2>
- [8] Shipra, B., Kibria, B.M.G. and Sharma, D. (2012) Testing the Population Coefficient of Variation. *Journal of Modern Applied Statistical Methods*, **11**, 325-335. <http://digitalcommons.wayne.edu/jmasm/vol11/iss2/5>
<https://doi.org/10.22237/jmasm/1351742640>
- [9] Zhaoyuan, L. and Jianfeng, Y. (2017) Testing for Heteroscedasticity in High-dimensional Regressions. Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong. <https://arxiv.org/abs/1510.00097>
<https://doi.org/10.1016/j.ecosta.2018.01.001>
- [10] Crépon, B. (2003) Économétrie linéaire. <http://www.crest.fr/ckfinder/userfiles/files/Pageperso/crepon/poly20052006.pdf>
- [11] Chekroun, A. (2017) Statistiques descriptives et exercices. Universiti Abou Bekr Belkaid Tlemcen-Algérie. <https://www.coursehero.com/file/27893347/chekroun-statistiquespdf/>
- [12] Bertoneche, M. (1979) Existence d'hétéroscédasticité dans le modèle de marché appliqué aux bourses européennes de valeurs mobilières. *Journal de la société statistique de Paris*, tome 120, 270-276. http://www.numdam.org/item/JFSFS_1979__120_4_270_0
- [13] Gastwirth, J.L., Gel, Y.R. and Miao, W. (2009) The Impact of Levene's Test of Equality of Variances on Statistical Theory and Practice. *Statistical Science*, **24**, 343-360. <https://doi.org/10.1214/09-STS301>

- [14] Vessereau, A. (1974) Essais interlaboratoires pour l'estimation de la fidélité des méthodes d'essais. *Revue de statistique appliquée*, tome 22, 5-48.
http://www.numdam.org/item/RSA_1974__22_1_5_0/