A computational approach test for comparing two linear regression models with unequal variances

Mehmet Enes Yazıcı[1] ID, Fikri Gökpınar*[2] ID, Esra Gökpınar[2] ID, Meral Ebegil[2] ID, Yaprak Arzu Özdemir[2] ID

[1] *Social Security Institution, Ankara, Turkey*
[2] *Department of Statistics, Faculty of Science, Gazi University, Ankara, Turkey*

Abstract

In this study, a new testing procedure is proposed to compare two linear regression models based on a computational approach test when the variances are not assumed equal. This method is based on restricted maximum likelihood estimators and some simple computational steps. To assess performance of the proposed test, it was compared with some existing tests in terms of power and type I error rate of the test. The simulation study reveals that the proposed test is a better alternative than some existing tests in most cases considered. Besides, an illustration of the proposed test was given by using a sample dataset.

Mathematics Subject Classification (2020). 62F03, 65C05, 65C60

Keywords. Chow test, computational approach test, parametric bootstrap test, heteroscedasticity regression models

1. Introduction

In many statistical applications, one of the most popular problems is the equality of the parameters of several independent populations or models fitted on these populations. There are many studies on this topic such as comparison of proportions, means, variance, stationary processes, and regression models, some of which are [14, 20–22, 27].

Regression models are widely applied to the measurement of economic relationships in econometrics where data are usually collected over a period. For this reason, it may be desirable to reveal whether the same relationship remains stable or not in different periods. For example, economic relations may be examined during pre-world war II and post-world war II to find out if they remain the same or not. Sometimes models developed in one period may lose their effectiveness as conditions change. The regression model, which is established before the economic crisis, cannot reflect the conditions after the economic crisis. Thus, a regression model before the economic crisis can change over time and can no longer be valid. Thus, it should be known whether it remains the same or not.

Regression models are also applied in many fields, such as medical studies, sport branches, etc. Researchers in medical studies may want to determine whether there is a significant difference between the regression models based on the effect of smoking on life expectancy for male and female populations. This difference can emerge by testing the equality of two linear regression models using statistical analysis.

Chow [5] proposed a test statistic called classical F test to test the equality of two linear regression models under the assumption of homogeneity of error variances. However, this assumption may unlikely be satisfied in many real-life problems. If the variances are not equal, the type I error probability of Chow test is negatively affected and gets higher than its nominal level. Consequently, testing the equality of regression models by using Chow test under heteroscedasticity may give quite inaccurate results. For this reason, when the assumption of homogeneity is not valid, several modifications of Chow test and new testing methods have been proposed in the literature. Toyoda [33] investigated the accuracy of the Chow test under heteroscedasticity and proposed a new test by altering the rejection region of the Chow test. Jayatissa [17] established a test for testing the equality of two linear regression models under heteroscedasticity. Conerly and Mansfield [6] suggested an approximate test by modifying the Chow statistic. Watt [34] proposed another alternative test called the Wald test and compared it numerically with Jayatissas test. He found out that the Wald test was powerful but had unreliable size for small samples. Schmidt and Sickles [29] examined Toyoda's approximation and found that the type I error rates of this approximation were not close to the nominal level especially when the two sample sizes and the two variances were very different. One of the most widely used tests, called weighted F test, was proposed by Goldfeld and Quandt [12], Kadiyala and Gupta [18], and Gupta [13]. However, this test does not work well for small samples. Ali and Silver [1] proposed two approximate tests to compare regression models based on usual F test and likelihood ratio test for the unequal variance case. Thursby [31] compared several exact and approximate tests for the equality of two linear regression models under heteroscedasticity. Moreno et al. [23] proposed a Bayesian solution to the problem of testing the equality of two linear regression models when the error variances were unknown and arbitrary. Oberhelman and Kadiyala [24] made some modifications of the standard F-test for large samples commonly used to test the equality of two linear regression models when the variances of the regression errors are not equal. Recently, there are many studies for testing the equality of regression models in several heteroscedasticity normal regression models [28, 32, 35].

The bootstrap procedure can be used to achieve higher-order- accurate inference in hypothesis testing problem. Third-order accuracy is obtained by inference procedures based on estimation of the sampling distribution of an appropriate statistic under the model in which the nuisance parameters are defined as their restricted maximum likelihood values for the given value of the interest parameter, see [8]. The Bootstrap procedure can also be applied to test the equality of two linear regression models. Tian et al. [32] offered a testing procedure based on Parametric Bootstrap (PB) approach to test the equality of regression models in several heteroscedasticity normal regression models. The Computational Approach Test (CAT) is an alternative PB method and based on restricted maximum likelihood estimation under null hypothesis. One of the advantages of this procedure is that it does not require the knowledge of any sampling distribution.

The aim of this paper is to obtain a procedure based on the CAT method to test the equality of two linear regression models under heteroscedasticity. The CAT method was firstly introduced by [26]. The CAT offers a proper solution especially when the type I error probability of traditional approaches is not close the nominal level. As mentioned in the references given above, type I error probabilities of most of the traditional tests are not close to the nominal level in many cases. For this reason, the CAT method can be very useful in this problem to adjust the type I error probability, especially in small

sample sizes. Some studies related to the CAT are available in the literature. Pal et al. [26] applied the CAT to Gamma and Weibull distributions for hypothesis testing and interval estimations. Chang and Pal [3] also used the CAT for testing the equality of two normal population means under heteroscedasticity. For Poisson and Gamma models, a testing procedure based on the CAT was given by [2, 4]. To test the equality of normal means under heteroscedasticity, a new testing procedure using the CAT was given by [9]. In another study, Gökpınar et al. [10] also proposed the CAT for the equality of several inverse Gaussian means under heterogeneity of scale parameters. In addition, Gökpınar and Gökpınar [11] used the CAT to test the equality of several log normal means. Since the CAT procedure is a good alternative for various testing procedures in these studies, the CAT method was applied to the equality of the two regression models under heteroscedasticity. Although it is easy to know the distribution of test statistic for the equality of two linear regression models under the assumption of homogeneity of error variances, in many real problems, this assumption is not satisfied. For this reason, in this study, we focus on a test statistic for the equality of two linear regression models when this assumption is not satisfied. We propose a test statistic based on CAT approach. This test has several advantageous especially for small sample size since the p-value of this proposed test statistic is obtained by CAT approach.

The rest of this study was organized as follows. In Section 2, the Chow test, Toyoda test, Wald test, weighted F test, and PB test were briefly introduced. In Section 3, the concept of the CAT procedure and a test based on this procedure to compare two linear regression models when the variances are not assumed equal was given. In Section 4, a simulation study was presented to assess the type I error rates and powers of the proposed test and the tests mentioned above. Furthermore, a numerical example was given in Section 5 and concluding remarks were summarized in Section 6.

2. Test statistics

In this section, some testing methods were given to test the equality of the two linear regression models under heteroscedasticity. Let two independent regression models with size n_1 and n_2 be as follows, respectively.

$$Y_i = X_i\beta_i + \varepsilon_i, \quad i = 1, 2 \tag{2.1}$$

Here, Y_i is a $n_i \times 1$ vector of observations, X_i is a known $n_i \times p$ design matrix of rank p, β_i is a $p \times 1$ vector of parameters and ϵ_i is a $n_i \times 1$ vector of error terms.

Error terms ϵ_i $(i = 1, 2)$ are distributed as normal with zero means and σ_i^2 $(i = 1, 2)$, which are arbitrary. To test the equality of the parameter vectors, the following hypotheses were given as

$$H_0: \ \beta_1 = \beta_2 \qquad H_1: \ \beta_1 \neq \beta_2. \tag{2.2}$$

Under H_0, the following function can be written as

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}\beta + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} = X\beta + \varepsilon, \tag{2.3}$$

where $\beta_1 = \beta_2 = \beta$, $\epsilon \sim (0, \sum)$, $\sum = \begin{pmatrix} \sigma_1^2 I_{n_1} & 0 \\ 0 & \sigma_2^2 I_{n_2} \end{pmatrix}$.

Residual sum of the squares of the model which was given in Eq.2.3 can be defined as follows:

$$
\begin{aligned}
e'e &= Y' \left[I - X \left(X'X \right)^{-1} X' \right] Y \\
&= \varepsilon' \left[I - X \left(X'X \right)^{-1} X' \right] \varepsilon \\
&= \varepsilon' \left[I - P_X \right] \varepsilon,
\end{aligned}
$$

where $P_X = X (X'X)^{-1} X'$ is a hat matrix. Under H_1, the residual sum of the squares can also be defined as follows:

$$e_i' e_i = Y_i' \left[I - X_i (X_i' X_i)^{-1} X_i' \right] Y_i = Y_i' \left[I - P_{X_i} \right] Y_i, \qquad i = 1, 2$$

where $P_{X_i} = X_i (X_i' X_i)^{-1} X_i', \; i = 1, 2$ are hat matrices for each model. Unrestricted residual sum of the squares can be obtained as follows:

$$e_1' e_1 + e_2' e_2 = Y_1' \left[I - P_{X_1} \right] Y_1 + Y_2' \left[I - P_{X_2} \right] Y_2 = Y' \left[I - P_{X^*} \right] Y,$$

where $X^* = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix}$ and $P_{X^*} = \begin{pmatrix} P_{X_1} & 0 \\ 0 & P_{X_2} \end{pmatrix}$. Since $[I - P_{X^*}] X^* = 0$ unrestricted residual sum of the squares can be written as

$$e_1' e_1 + e_2' e_2 = \varepsilon' \left[I - P_{X^*} \right] \varepsilon.$$

In the rest of this section, the most widely used test statistics for this problem are presented.

2.1. Chow test

To test the equality of two linear regression models under homogeneity of error variances, the Chow test statistic can be obtained as follows:

$$F = \frac{[e'e - e_1' e_1 - e_2' e_2]/p}{[e_1' e_1 + e_2' e_2]/(n_1 + n_2 - 2p)} = \frac{\varepsilon' [P_{X^*} - P_X] \varepsilon/p}{\varepsilon' [1 - P_{X^*}] \varepsilon/(n_1 + n_2 - 2p)} \tag{2.4}$$

The test statistic given in Eq.2.4 has F distribution with degrees of freedom p and $(n_1 + n_2 - 2p)$ respectively when $\sigma_1^2 = \sigma_2^2$ under H_0 [5].

Toyoda [33] showed that the Chow test is well behaved even under heteroscedasticity as long as at least one of the two sample sizes is very large. However, for two small sample sizes, the type I error rates of this test is affected considerably from even moderate heteroscedasticity.

2.2. Toyoda test

Toyoda test was obtained using Satterhwaite approximation as follows:

$$F^* = \frac{[e'e - (e_1' e_1 + e_2' e_2)]/p}{(e_1' e_1 + e_2' e_2)/f_2} = \frac{\varepsilon' [P_{X^*} - P_X] \varepsilon/p}{\varepsilon' [1 - P_{X^*}] \varepsilon/(n_1 + n_2 - 2p)}, \tag{2.5}$$

where $f_2 = \frac{\left[(n_1-p)\hat{\sigma}_1^2 + (n_2-p)\hat{\sigma}_2^2 \right]^2}{(n_1-p)\hat{\sigma}_1^4 + (n_2-p)\hat{\sigma}_2^4}$. F^* statistic is distributed as F with degrees of freedom p and f_2 [33]. Schmidt and Sickles [29] found out that Toyodas approximation has a reasonable accuracy only when the two variances are of the same order of magnitude and the sample sizes are also of the same order of magnitude.

2.3. Wald test

The Wald test statistic was initially suggested by [17, 34]. This statistic is defined as follows:

$$W = \left(\hat{\beta}_1 - \hat{\beta}_2 \right)' \left(\hat{\sigma}_1^2 (X_1' X_1)^{-1} + \hat{\sigma}_2^2 (X_2' X_2)^{-1} \right)^{-1} \left(\hat{\beta}_1 - \hat{\beta}_2 \right), \tag{2.6}$$

where $\hat{\beta}_i = (X_i' X_i)^{-1} X_i' Y_i$ and $\hat{\sigma}_i^2 = \frac{(Y_i - X_i \hat{\beta}_i)'(Y_i - X_i \hat{\beta}_i)}{(n_i - p)}$ are least squares estimators of β_i and σ_i^2 $(i = 1, 2)$, respectively. The Wald test statistic is approximately distributed as χ_p^2. Watt [34] compared Jayatissa [17]'s test numerically with the Wald test and found

out that the Wald test was powerful for large samples but has unreliable size for small samples. This result was confirmed both theoretically and numerically by [16].

2.4. Weighted F test

Weighted F test is a modification of the Wald test and can be given as follows:

$$F_w = \frac{W}{p} \tag{2.7}$$

This test statistic is distributed as F with degrees of freedom p and $n_1 + n_2 - 2p$ [13].

2.5. Parametric bootstrap test

Tian et al. [32] provided the PB method for testing Eq. 2.2. In this paper, the algorithm of this method was given for only two groups.

Step 1: Let $S_i^2 = \hat{\sigma}_i^2$ and the observed value t of T given below is calculated as

$$
\begin{aligned}
T &= \sum_{i=1}^2 S_i^{-2} \left(\widehat{\beta}_i - \hat{\beta}\right)' (X_i'X_i) \left(\widehat{\beta}_i - \hat{\beta}\right) \\
&= \sum_{i=1}^2 S_i^{-2} \widehat{\beta}_i' (X_i'X_i) \widehat{\beta}_i - \hat{\beta}' \left(\sum_{i=1}^2 S_i^{-2} (X_i'X_i)\right) \hat{\beta} \\
&= \sum_{i=1}^2 S_i^{-2} X_i \widehat{\beta_i^2} - \sum_{i=1}^2 S_i^{-2} X_i \hat{\beta}^2.
\end{aligned} \tag{2.8}
$$

Step 2: Independent pseudo numbers are generated from $Z_i = (Z_{i1}, \ldots, Z_{ip}) \sim MN(0, I_p)$ and $U_i \sim \chi^2_{(n_i-p)}, \quad i = 1, 2.$

Step 3: T_{PB} value is calculated as follows:

$$
\begin{aligned}
T_{PB} = &\sum_{i=1}^2 s^{-2} Z_i' Z_i \\
&- \left(\sum_{i=1}^2 \frac{(n_i-p)}{s_i U_i} (X_i'X_i)^{1/2} Z_i\right)' \left(\sum_{i=1}^2 \frac{(n_i-p)}{s_i^2 U_i} (X_i'X_i)\right)^{-1} \left(\sum_{i=1}^2 \frac{(n_i-p)}{s_i U_i} (X_i'X_i)^{1/2} Z_i\right)
\end{aligned}
$$

Step 4: Step 2-3 are repeated many times, say m, and estimated *p*-value are calculated as follows:

$$\hat{p} = \sum_{j=1}^m \frac{\left(T_{PB}^j \geqslant t\right)}{m}$$

For a given level α, the PB test rejects the null hypothesis when the $\hat{p}$ is less than α.

3. The computational approach test

To understand the CAT better, we first presented the general framework of this method. Then, we presented the newly proposed testing method based on the CAT to test the equality of two linear regression models under heteroscedasticity.

Let $U_1, U_2, \ldots, U_n$ be a random sample having a probability density function $f(u/\theta)$, where the functional form of f is assumed to be known and $\theta = \left(\theta^{(1)}, \theta^{(2)}\right)$ is an unknown vector in parameter space Θ. $\theta^{(1)}$ and $\theta^{(2)}$ are the parameter of interest and nuisance parameter, respectively. The problem of interest is to test $H_0' : \theta^{(1)} = \theta_0^{(1)}$ versus a suitable alternative. To test $H_0' : \theta^{(1)} = \theta_0^{(1)}$ against H_1', H_0' was first expressed as

$H_0'^* : \eta\left(\theta^{(1)}, \theta_0^{(1)}\right) = 0$ against $H_1'^*$, where η is a scalar valued function. The general methodology of the CAT for testing $H_0'^* : \eta\left(\theta^{(1)}, \theta_0^{(1)}\right) = 0$ against a suitable alternative at a desired level α was given through the following steps [26].

(1) Obtain maximum likelihood estimations (MLEs) of the parameters $\theta^{(1)}$ and $\theta^{(2)}$.

(2) Obtain a suitable $\eta\left(\theta^{(1)}, \theta_0^{(1)}\right)$ and the MLEs of η, $\hat{\eta} = \hat{\eta}\left(\hat{\theta}^{(1)}, \theta_0^{(1)}\right)$ can be used as a test statistic.

(3) Under Ho, find the MLEs of $\theta^{(2)}$ parameter, which is denoted by $\widetilde{\theta}^{(2)}$.

(4) Generate artificial sample $U_1, U_2, \ldots, U_n$ from $f\left(u/\theta_0^{(1)}, \widetilde{\theta}^{(2)}\right)$ arge number of times, say m times. For each of these replicated samples, recalculate the MLE η, $\hat{\eta}^{(j)}$ where $j = 1, \ldots, m$.

(5) Estimate the p-value as $\hat{p} = \sum_{j=1}^{m} \frac{\left(\hat{\eta}^{(j)} \geqslant \hat{\eta}\right)}{m}$. In the case of $\hat{p} < \alpha$, H_0 is rejected.

There are two important points to obtain this algorithm; the first of which is to obtain a suitable η and the second one is to obtain RMLEs of the parameters under H_0. Accordingly, before we carry out the CAT to test the equality of two linear regression models under heteroscedasticity, we first need to find the RMLEs of the model parameters as given in Theorem 3.1.

Theorem 3.1. *The RMLEs of the model parameters β, σ_1^2 and σ_2^2 were given as follows:*

$$\widetilde{\beta} = \left[\frac{\left(\sum_{i=1}^{n_1} X_{1i} X_{1i}'\right)}{\widetilde{\sigma}_1^2} + \frac{\left(\sum_{i=1}^{n_2} X_{2i} X_{2i}'\right)}{\widetilde{\sigma}_2^2}\right]^{-1} \left(\frac{\sum_{i=1}^{n_1} Y_{1i} X_{1i}'}{\widetilde{\sigma}_1^2} + \frac{\sum_{i=1}^{n_2} Y_{2i} X_{2i}'}{\widetilde{\sigma}_2^2}\right)$$

$$\widetilde{\sigma}_1^2 = \frac{1}{n_1}\left[\sum_{i=1}^{n_1} \left(Y_{1i} - X_{1i}'\widetilde{\beta}\right)^2\right]$$

$$\widetilde{\sigma}_2^2 = \frac{1}{n_2}\left[\sum_{i=1}^{n_2} \left(Y_{2i} - X_{2i}'\widetilde{\beta}\right)^2\right]$$

Proof. The restricted log-likelihood function under H_0 can be obtained as follows:

$$\ln(L) = -\frac{n_1 + n_2}{2}\ln(2\pi) - \frac{n_1}{2}\ln\left(\sigma_1^2\right) - \frac{n_2}{2}\ln\left(\sigma_2^2\right)$$
$$- \frac{1}{2\sigma_1^2}\left[\sum_{i=1}^{n_1} Y_{1i}^2 - 2\beta\sum_{i=1}^{n_1} Y_{1i}X_{1i}' + \beta'\left(\sum_{i=1}^{n_1} X_{1i}X_{1i}'\right)\beta\right]$$
$$- \frac{1}{2\sigma_2^2}\left[\sum_{i=1}^{n_2} Y_{2i}^2 - 2\beta\sum_{i=1}^{n_2} Y_{2i}X_{2i}' + \beta'\left(\sum_{i=1}^{n_2} X_{2i}X_{2i}'\right)\beta\right]$$

The score functions of the parameters β, σ_1^2 and σ_2^2 are as follows:

$$\frac{\ln(L)}{\partial\beta} = S(\beta) = -\frac{1}{2\sigma_1^2}\left[-2\sum_{i=1}^{n_1} Y_{1i}X_{1i}' + 2\beta'\sum_{i=1}^{n_1} X_{1i}X_{1i}'\right]$$
$$- \frac{1}{2\sigma_2^2}\left[-2\sum_{i=1}^{n_2} Y_{2i}X_{2i}' + 2\beta'\sum_{i=1}^{n_2} X_{2i}X_{2i}'\right] \tag{3.1}$$

$$\frac{\ln(L)}{\partial\sigma_1^2} = S\left(\sigma_1^2\right) = -\frac{n_1}{2\left(\sigma_1^2\right)} + \frac{1}{2\left(\sigma_1^2\right)^2}\left[\sum_{i=1}^{n_1} \left(Y_{1i} - X_{1i}'\beta\right)^2\right] \tag{3.2}$$

$$\frac{\ln(L)}{\partial\sigma_2^2} = S\left(\sigma_2^2\right) = -\frac{n_2}{2\left(\sigma_2^2\right)} + \frac{1}{2\left(\sigma_2^2\right)^2}\left[\sum_{i=1}^{n_2} \left(Y_{2i} - X_{2i}'\beta\right)^2\right] \tag{3.3}$$

Using these score functions, the RMLEs of the parameters can be obtained as follows:

$$\widetilde{\beta} = \left[\frac{(\sum_{i=1}^{n_1} X_{1i} X'_{1i})}{\widetilde{\sigma}_1^2} + \frac{(\sum_{i=1}^{n_2} X_{2i} X'_{2i})}{\widetilde{\sigma}_2^2} \right]^{-1} \left(\frac{\sum_{i=1}^{n_1} Y_{1i} X'_{1i}}{\widetilde{\sigma}_1^2} + \frac{\sum_{i=1}^{n_2} Y_{2i} X'_{2i}}{\widetilde{\sigma}_2^2} \right)$$

$$\widetilde{\sigma}_1^2 = \frac{1}{n_1} \left[\sum_{i=1}^{n_1} \left(Y_{1i} - X'_{1i} \widetilde{\beta} \right)^2 \right]$$

$$\widetilde{\sigma}_2^2 = \frac{1}{n_2} \left[\sum_{i=1}^{n_2} \left(Y_{2i} - X'_{2i} \widetilde{\beta} \right)^2 \right]$$

$\square$

As seen in Theorem 3.1, there are no close forms of the RMLEs of the parameters. For this reason, some iterative algorithms should be used to obtain the RMLEs of the parameters. In this study, Fisher-Scoring algorithm was used to obtain these estimators. Fisher scoring algorithm is a modified version of Newton-Raphson method, especially to obtain the MLEs of the parameters. This algorithm uses the expected information matrix, while Newton-Raphson algorithm uses the observed information matrix. This makes Fisher scoring less dependent on specific data values. Knight [19] and Schworer and Hovey [30] pointed out the differences between the Newton-Raphson and Fisher scoring algorithms:

(1) The convergence of the Newton-Raphson algorithm is often faster when both algorithms converge, while Fisher scoring algorithm is more robust and will converge when Newton-Raphson algorithm does not.

(2) The radius of the convergence for the Fisher scoring algorithm is often larger which suggests that the choice of an initial estimate is less important for the Fisher scoring algorithm. For these reasons, we used Fisher scoring algorithm to obtain the RMLEs. The ith iteration of this algorithm can be given as follows:

$$\theta_{i+1} = \theta_i + \gamma \left\{ I(\theta) + S(\theta) S(\theta)' \right\}^{-1} S(\theta), \quad i = 0, 1, 2, \dots$$

where $0 < \gamma < 1$ is an arbitrary constant, $S(\theta)$ is the score vector of the parameters, and $I(\theta)$ is the expected value of the information matrix, $J(\theta)$. In this problem, the parameter vector θ is $(\beta, \sigma_1^2, \sigma_2^2)$. The elements of the score vector are given at Eq.3.1, Eq. 3.2 and Eq. 3.3. The information matrix and its expectation of this model are also needed to use this algorithm. The information matrix and its expectation were given as follows:

$$J(\beta, \sigma_1^2, \sigma_2^2) = \begin{pmatrix} a & b_1 & b_2 \\ 0 & c_1 & 0 \\ 0 & 0 & c_2 \end{pmatrix},$$

where

$$a = \frac{\sum_{i=1}^{n_1} X_{1i} X'_{1i}}{\sigma_1^2} + \frac{\sum_{i=1}^{n_2} X_{2i} X'_{2i}}{\sigma_2^2},$$

$$b_1 = \frac{\sum_{i=1}^{n_1} Y_{1i} X'_{1i}}{(\sigma_1^2)^2} - \frac{\beta' (\sum_{i=1}^{n_1} X_{1i} X'_{1i})}{(\sigma_1^2)^2},$$

$$b_2 = \frac{\sum_{i=1}^{n_2} Y_{2i} X'_{2i}}{(\sigma_2^2)^2} - \frac{\beta' (\sum_{i=1}^{n_2} X_{2i} X'_{2i})}{(\sigma_2^2)^2},$$

$$c_1 = -\frac{n_1}{2(\sigma_1^2)} + \frac{1}{2(\sigma_1^2)^3} \left[\sum_{i=1}^{n_1} (Y_{1i} - X'_{1i}\beta)^2 \right],$$

$$c_2 = -\frac{n_2}{2(\sigma_2^2)} + \frac{1}{2(\sigma_2^2)^3} \left[\sum_{i=1}^{n_2} (Y_{2i} - X'_{2i}\beta)^2 \right],$$

and

$$I\left(\beta,\ \sigma_1^2,\sigma_2^2\right) = E\left(J\left(\beta,\ \sigma_1^2,\sigma_2^2\right)\right) = \begin{pmatrix} \frac{\sum_{i=1}^{n_1} X_{1i}X'_{1i}}{\sigma_1^2} + \frac{\sum_{i=1}^{n_2} X_{2i}X'_{2i}}{\sigma_2^2} & 0 & 0 \\ 0 & \frac{n_1-2p}{2\left(\sigma_1^2\right)^2} & 0 \\ 0 & 0 & \frac{n_2-2p}{2\left(\sigma_2^2\right)^2} \end{pmatrix}$$

(3.4)

The score vector of $\theta = (\beta,\sigma_1^2,\sigma_2^2)$ can also be rewritten as follows:

$$S\left(\beta,\sigma_1^2,\sigma_2^2\right) = \left(S\left(\beta\right),S\left(\sigma_1^2\right),S\left(\sigma_1^2\right)\right),$$

(3.5)

where

$$S\left(\beta\right) = \sum_{j=1}^2 \frac{1}{\sigma_j^2}\sum_{i=1}^{n_j} Y_{ji}X'_{ji} - \beta'\sum_{j=1}^2\sum_{i=1}^{n_j} X_{ji}X'_{ji},$$

$$S\left(\sigma_1^2\right) = -\frac{n_1}{2\left(\sigma_1^2\right)} + \frac{1}{2\left(\sigma_1^2\right)^2}\left[\sum_{i=1}^{n_1}\left(Y_{1i} - X'_{1i}\beta\right)^2\right],$$

$$S\left(\sigma_1^2\right) = -\frac{n_2}{2\left(\sigma_2^2\right)} + \frac{1}{2\left(\sigma_2^2\right)^2}\left[\sum_{i=1}^{n_2}\left(Y_{2i} - X'_{2i}\beta\right)^2\right].$$

For the initial value of θ, θ_0 was taken as

$$\theta_0 = \left(\widetilde{\beta}_{(0)},\widetilde{\sigma}^2_{1(0)},\widetilde{\sigma}^2_{2(0)}\right)$$

$$= \left\{(X'X)^{-1}X'Y,\ \frac{1}{n_1-p}\left[\sum_{i=1}^{n_1}\left(Y_{1i} - X'_{1i}\widetilde{\beta}_{(0)}\right)^2\right],\ \frac{1}{n_2-p}\sum_{i=1}^{n_2}\left(Y_{2i} - X'_{2i}\widetilde{\beta}_{(0)}\right)^2\right\}.$$

(3.6)

Remark 3.2. We observed that when γ s close to the upper bound, the length of the steps from θ_i to θ_{i+1} can be great. Thus, when the initial value of θ_0 is far from the optimal value, it might wildly oscillate and not converge at all. For this reason, this case is sometimes remedied by making smaller steps. However, the length of the steps from θ_i to θ_{i+1} can be smaller by taking γ too close to 0 and this causes the optimal value to converge very slowly. When we took $\gamma = 0.1$ or this case, we observed in our simulation study that the Fisher Scoring algorithm converges in a few iterations with the initial value defined above. This algorithm can be summarized as given:

Algorithm 1: (Fisher-Scoring Algorithm for the RMLEs of the parameters for the equality of two linear regression models):

(1) Obtain the initial value of the maximum likelihood estimation as given in Eq. 3.6:

$$\theta_0 = \left(\widetilde{\beta}_{(0)}\widetilde{\sigma}^2_{1(0)},\widetilde{\sigma}^2_{2(0)}\right)$$

$$\widetilde{\beta}_{(0)} = (X'X)^{-1}X'Y)$$

$$\hat{\sigma}^2_{1(0)} = \frac{1}{n_1-p}\left[\sum_{i=1}^{n_1}\left(Y_{1i} - X'_{1i}\widetilde{\beta}_0\right)^2\right]$$

$$\hat{\sigma}^2_{2(0)} = \frac{1}{n_2-p}\sum_{i=1}^{n_2}\left(Y_{2i} - X'_{2i}\widetilde{\beta}_0\right)^2$$

(2) Using Step 1, obtain $I\left(\theta_0\right)$ and $S\left(\theta_0\right)$ as given in Eq. 3.4 and Eq.3.5.
(3) Using Eq. 3.4, obtain $\theta_{i+1},\ i = 1,2,3,\ldots$.
(4) Repeat the Steps (1-3) until $|\theta_{i+1} - \theta_i| < \varepsilon$.
(5) The RMLEs of the parameters $\beta,\sigma_1^2,\sigma_2^2$ under H_0 are $\theta_{i+1} = \left(\widetilde{\beta}_{(i+1)},\ \widetilde{\sigma}^2_{1(i+1)},\widetilde{\sigma}^2_{2(i+1)}\right)$.

Remark 3.3. As previously mentioned, one of the two important points to apply the CAT method is to obtain a suitable test statistic denoted as η. Davison and Hinkley [7] expressed that the likelihood-based methods (the Wald, likelihood ratio and Score tests) are the most appropriate tests for the PB method. The Wald, likelihood ratio, and Score tests are all asymptotically optimal and provide good properties in large samples. Although these tests are equivalent in large samples, they are different in small samples. Therefore, it is important for a test to not only perform well only on large samples, but also well in small samples. Ohtani and Toyoda [25] compared the size-corrected Wald, likelihood ratio, and Score tests by Monte Carlo experiments to test the equality of two linear regression models under heteroscedasticity in small sample sizes. Their study showed that the size-corrected Wald and Score tests perform quite well. However, it is known that the Wald test depends only on MLEs of parameters while the Score test depends on both MLEs and restricted MLEs under the null hypothesis. Thus, the calculation of the Score test is quite complicated. Besides, the size correction of a test depends on the nominal level of the type I error. Recomputation of size correction of a test is required for each different nominal level of the type I error. However, this case is not practical in real world problems. At this point, the CAT procedure can be very useful for this kind of problems adjusting the type I error rates of any test automatically. For this reason, we carried out the CAT to test the equality of two linear regression models under heteroscedasticity using the Wald test statistic in Eq. 2.6.

The CAT for the equality of two linear regression models can be given as:

Algorithm 2: (the CAT for the equality of two linear regression models):

(1) Obtain the unrestricted MLE for $\hat{\theta} = \left(\hat{\beta}_1, \ \hat{\beta}_2, \ \hat{\sigma}_1^2, \hat{\sigma}_2^2 \right)$ as follows:

$$\hat{\beta}_1 = \left(X_1' X_1 \right)^{-1} X_1' Y_1, \quad \hat{\beta}_2 = \left(X_2' X_2 \right)^{-1} X_2' Y_2,$$

$$\hat{\sigma}_1^2 = \frac{1}{n_1 - p} \left[\sum_{i=1}^{n_1} \left(Y_{1i} - X_{1i}' \hat{\beta}_1 \right)^2 \right], \ \text{ and } \ \hat{\sigma}_2^2 = \frac{1}{n_2 - p} \left[\sum_{i=1}^{n_2} \left(Y_{2i} - X_{2i}' \hat{\beta}_2 \right)^2 \right].$$

(2) Obtain the test statistic $\hat{\eta}$ as follows:

$$\hat{\eta} = \left(\hat{\beta}_1 - \hat{\beta}_2 \right)' \left(\hat{\sigma}_1^2 \left(X_1' X_1 \right)^{-1} + \hat{\sigma}_2^2 \left(X_2' X_2 \right)^{-1} \right)^{-1} \left(\hat{\beta}_1 - \hat{\beta}_2 \right).$$

(3) Obtain RMLE for $\widetilde{\theta} = \left(\widetilde{\beta}, \ \widetilde{\sigma}_1^2, \widetilde{\sigma}_2^2 \right)$ using Algorithm 1 under H_0.

(4) Generate artificial sample using RMLEs.

$$\hat{Y}_1 = \widetilde{\beta} X_1 + \varepsilon_1^*, \quad \varepsilon_1^* \sim N \left(0, \widetilde{\sigma}_1^2 \right)$$

$$\hat{Y}_2 = \widetilde{\beta} X_2 + \varepsilon_2^*, \quad \varepsilon_2^* \sim N \left(0, \widetilde{\sigma}_2^2 \right)$$

(5) Calculate $\widetilde{\eta}$ for the generated artificial sample given in Step 4.

(6) Repeat Step 4-5 for large number of times, say m times.

(7) Estimate the p-value as $\hat{p} = \sum_{j=1}^m \frac{\left(\hat{\eta}^{(j)} \geqslant \widetilde{\eta} \right)}{m}$. In the case of $\hat{p} < \alpha$, H_0 is rejected.

Remark 3.4. The CAT is a PB method since resampling is generated from the parametric model $\varepsilon_i^* \sim N \left(0, \widetilde{\sigma}_i^2 \right), \ i = 1, 2$. Artificial sample is generated to mimic the null distribution of $\hat{\eta}$. Thus, the cut-off point of $\hat{\eta}$ is an approximation of the true critical value based on the null model, under H_0. The CAT does not require any knowledge of the complex sampling distribution of a test statistic and finds the critical region automatically. The proposed CAT heavily depends on computations and simulations, but with the computational resources available today, this method can be used effectively. As a result, the CAT can be said to be easy to implement and compute.

4. Simulation study

In this section, the proposed test (CAT) was compared with the Chow (C) test, Toyoda (T) test, Wald (W) test, weighted F (WF) test, and PB test in terms of their estimated type-I error rates and powers of test. For this purpose, the explanatory variables $p=1$, 2, and 3 were considered as different combinations of equal and unequal sample sizes $(n = (n_1, n_2))$. The standard deviations of the regression models $(\sigma = (\sigma_1, \sigma_2))$ equal to (1,4) and (1,8), respectively. To estimate the powers of tests, we took three different parameter combinations of regression coefficients. To estimate the type-I error rates and powers of all tests under the specified nominal level $\alpha=0.05$, 5000 random numbers were generated for errors with sample size n_i $(i = 1, 2)$ from the normal distribution. $m = 5000$ was also taken to estimate the p-values of the CAT and PB test. The simulation study was conducted in MATLAB. The estimated type-I error rates are presented in Table 1-Table 3.

Table 1. The type I error rates of the all tests when $p=1$.

n	$\sigma = (1, 4)$						$\sigma = (1, 8)$					
	C	T	W	WF	PB	CAT	C	T	W	WF	PB	CAT
5,5	0.077	0.031	0.150	0.111	0.087	0.047	0.075	0.022	0.159	0.115	0.086	0.047
10,10	0.060	0.018	0.095	0.076	0.064	0.047	0.063	0.014	0.096	0.078	0.063	0.046
15,15	0.055	0.016	0.074	0.064	0.057	0.048	0.053	0.012	0.072	0.062	0.051	0.044
25,25	0.061	0.015	0.069	0.063	0.058	0.052	0.057	0.012	0.067	0.060	0.054	0.048
50,50	0.053	0.012	0.057	0.055	0.052	0.048	0.055	0.007	0.057	0.054	0.051	0.048
5,10	0.019	0.010	0.096	0.074	0.070	0.048	0.013	0.003	0.098	0.074	0.067	0.049
5,15	0.007	0.003	0.082	0.065	0.063	0.052	0.005	0.002	0.083	0.067	0.065	0.053
10,15	0.031	0.010	0.084	0.070	0.064	0.052	0.028	0.007	0.079	0.068	0.062	0.050
15,25	0.019	0.005	0.067	0.060	0.057	0.051	0.016	0.004	0.068	0.062	0.057	0.052
15,50	0.002	0.001	0.058	0.054	0.054	0.050	0.000	0.000	0.061	0.056	0.056	0.052
10,5	0.180	0.072	0.165	0.134	0.093	0.052	0.209	0.045	0.171	0.142	0.087	0.047
25,5	0.353	0.153	0.169	0.153	0.088	0.049	0.413	0.086	0.165	0.152	0.080	0.047
15,10	0.121	0.031	0.102	0.084	0.067	0.051	0.120	0.022	0.096	0.081	0.065	0.049
25,15	0.135	0.029	0.081	0.072	0.059	0.048	0.133	0.017	0.078	0.069	0.055	0.045
50,15	0.259	0.067	0.086	0.081	0.064	0.054	0.284	0.040	0.084	0.080	0.064	0.054

As seen from Table 1, when sample sizes are equal and small, the type I error rates of the CAT are close to the nominal level α. However, the type I error rates of the others are far away than the nominal level α. The type I error rates of the T test are much smaller than the nominal level, whereas those of the others are much greater than the nominal level. As sample sizes increase, the type I error rates of all tests are close to the nominal level α, whereas the T test seems to have much lower type I error rates than the nominal level α, even for large sample sizes.

When the sample sizes and variances are proportional, the type I error rates of the CAT are close to the nominal level α regardless of sample sizes. While the type I error rates of the W, WF, and PB tests are greater than the nominal level α in the cases of small sample sizes, these values of tests are close to the nominal level α in the cases of large sample sizes. As for the C and T tests, the type I error rates of these tests are much smaller than the nominal level α in the cases of all sample sizes. When the sample sizes and variances are in reverse order, the CAT appears to have type I error rates closer to the nominal level for all sample sizes. The others are much greater than the nominal level. It is also seen that the type I error rates of all the considered tests are not affected by the increase in variances.

As seen from Table 2 and Table 3, as the value of p increase, except for the CAT, the type I error rates of the others somewhat increase. The results of $p=2$ and $p=3$ are the same as the results of $p=1$ in terms of interpretation.

Table 2. The type I error rates of the all tests when $p=2$.

	$\sigma = (\mathbf{1}, \mathbf{4})$						$\sigma = (\mathbf{1}, \mathbf{8})$					
n	**C**	**T**	**W**	**WF**	**PB**	**CAT**	**C**	**T**	**W**	**WF**	**PB**	**CAT**
5,5	0.087	0.038	0.233	0.131	0.092	0.052	0.094	0.028	0.232	0.132	0.084	0.044
10,10	0.069	0.018	0.121	0.088	0.071	0.043	0.081	0.014	0.137	0.101	0.074	0.053
15,15	0.061	0.017	0.091	0.070	0.060	0.044	0.069	0.011	0.098	0.081	0.064	0.050
25,25	0.062	0.011	0.078	0.066	0.060	0.053	0.063	0.006	0.074	0.065	0.057	0.049
50,50	0.059	0.009	0.066	0.061	0.056	0.053	0.053	0.003	0.059	0.055	0.049	0.044
5,10	0.017	0.008	0.133	0.079	0.072	0.057	0.014	0.006	0.131	0.082	0.074	0.053
5,15	0.004	0.002	0.113	0.069	0.063	0.059	0.002	0.001	0.098	0.068	0.064	0.051
10,15	0.028	0.008	0.098	0.073	0.068	0.049	0.025	0.005	0.101	0.073	0.062	0.048
15,25	0.015	0.003	0.071	0.058	0.056	0.044	0.014	0.002	0.076	0.060	0.054	0.050
15,50	0.001	0.000	0.062	0.054	0.054	0.052	0.002	0.000	0.066	0.059	0.059	0.056
10,5	0.256	0.114	0.244	0.189	0.112	0.046	0.309	0.066	0.257	0.198	0.086	0.044
25,5	0.539	0.237	0.256	0.230	0.101	0.050	0.633	0.136	0.249	0.224	0.083	0.046
15,10	0.149	0.033	0.124	0.095	0.067	0.044	0.165	0.022	0.127	0.098	0.066	0.049
25,15	0.161	0.029	0.091	0.078	0.054	0.046	0.191	0.023	0.100	0.086	0.064	0.055
50,15	0.367	0.076	0.100	0.092	0.067	0.052	0.418	0.033	0.098	0.088	0.057	0.045

Table 3. The type I error rates of the all tests when $p=3$.

	$\sigma = (\mathbf{1}, \mathbf{4})$						$\sigma = (\mathbf{1}, \mathbf{8})$					
n	**C**	**T**	**W**	**WF**	**PB**	**CAT**	**C**	**T**	**W**	**WF**	**PB**	**CAT**
5,5	0.087	0.038	0.233	0.131	0.092	0.052	0.094	0.028	0.232	0.132	0.084	0.044
10,10	0.069	0.018	0.121	0.088	0.071	0.043	0.081	0.014	0.137	0.101	0.074	0.053
15,15	0.061	0.017	0.091	0.070	0.060	0.044	0.069	0.011	0.098	0.081	0.064	0.050
25,25	0.062	0.011	0.078	0.066	0.060	0.053	0.063	0.006	0.074	0.065	0.057	0.049
50,50	0.059	0.009	0.066	0.061	0.056	0.053	0.053	0.003	0.059	0.055	0.049	0.044
5,10	0.017	0.008	0.133	0.079	0.072	0.057	0.014	0.006	0.131	0.082	0.074	0.053
5,15	0.004	0.002	0.113	0.069	0.063	0.059	0.002	0.001	0.098	0.068	0.064	0.051
10,15	0.028	0.008	0.098	0.073	0.068	0.049	0.025	0.005	0.101	0.073	0.062	0.048
15,25	0.015	0.003	0.071	0.058	0.056	0.044	0.014	0.002	0.076	0.060	0.054	0.050
15,50	0.001	0.000	0.062	0.054	0.054	0.052	0.002	0.000	0.066	0.059	0.059	0.056
10,5	0.256	0.114	0.244	0.189	0.112	0.046	0.309	0.066	0.257	0.198	0.086	0.044
25,5	0.539	0.237	0.256	0.230	0.101	0.050	0.633	0.136	0.249	0.224	0.083	0.046
15,10	0.149	0.033	0.124	0.095	0.067	0.044	0.165	0.022	0.127	0.098	0.066	0.049
25,15	0.161	0.029	0.091	0.078	0.054	0.046	0.191	0.023	0.100	0.086	0.064	0.055
50,15	0.367	0.076	0.100	0.092	0.067	0.052	0.418	0.033	0.098	0.088	0.057	0.045

As a result, the type I error rates of the CAT are close to the nominal level α regardless of sample sizes, variance configuration, and the relation between sample sizes and variances. When sample sizes and variances are proportional, the type I error rates of the C and T tests are much smaller than the nominal level α. On the other hand, those of the C and T tests are much greater when sample sizes and variances are in reverse order. That is, the C and T tests are negatively affected by this situation. The W, WF, and PB tests have the type I error rates closer to nominal level when the sample sizes and variances are proportional and sample sizes are large. On the other hand, when sample sizes and variances are in reverse order, the W, WF, and PB tests appear to have greater type I error rates than nominal level α. That is, these tests are negatively affected by this situation. The powers of the tests are given in Table 4-9.

While the powers of the tests were interpreted, the tests, which had greater type I error rates than %6 given in Table 1-Table 3, were disregarded. Thus, the powers of these tests were denoted by ****.

As can be seen from Table 4–Table 9, the CAT seems to be more powerful than the other tests having lower type I error rates than %6 especially in the case of small and equal sample sizes. The powers of the PB and WF tests appear to be a bit more powerful

than the CAT, in the case of equal and large sample sizes and when the sample sizes and variances are proportional for large sample sizes. On the other hand, when sample sizes and variances were in reverse order, these tests were disregarded since the WF and PB tests have greater type I error rates than nominal level α. Thus, the CAT seems to be more powerful than the other tests in this case, regardless of sample size. The powers of all the tests are negatively affected from the increases in values of variance and as seen from tables, the powers of all tests are smaller. The CAT appears to be more powerful than the others in this case. When sample sizes increase, the powers of the PB test are a bit more powerful than the CAT. The powers of all the tests are negatively affected by the increases in values of variances and those of the tests are smaller. As expected, the powers of all tests increase when the difference among β values increase. As the values of p increase, all the tests are negatively affected by this case, and the powers of the tests decrease. The interpretation obtained from the cases of higher value of p are the same as those of the cases of $p=1$.

Table 4. The power of rates of the all tests when $p=1$, $\beta_1 = 1$, $\beta_2 = 2$.

n	$\sigma = (1,4)$						$\sigma = (1,8)$					
	C	T	W	WF	PB	CAT	C	T	W	WF	PB	CAT
5,5	****	0.040	****	****	****	0.059	****	0.024	****	****	****	0.051
10,10	****	0.046	****	****	****	0.100	****	0.016	****	****	****	0.058
15,15	0.146	0.046	****	****	0.156	0.133	0.078	0.014	****	****	0.079	0.067
25,25	0.218	0.085	****	0.236	0.229	0.213	0.096	0.023	****	****	0.098	0.087
50,50	0.382	0.179	0.407	0.398	0.392	0.383	0.138	0.034	0.151	0.144	0.139	0.136
5,10	0.034	0.017	****	****	****	0.093	0.019	0.007	****	****	****	0.059
5,15	0.017	0.007	****	****	****	0.125	0.009	0.004	****	****	****	0.069
10,15	0.078	0.030	****	****	0.157	0.135	0.039	0.009	****	****	****	0.075
15,25	0.100	0.044	0.245	0.228	0.222	0.204	0.032	0.009	****	****	0.106	0.097
15,50	0.055	0.028	0.381	0.367	0.364	0.348	0.007	0.002	****	0.151	0.150	0.142
10,50	****	****	****	****	****	0.067	****	0.052	****	****	****	0.057
25,50	****	****	****	****	****	0.068	****	****	****	****	****	0.050
15,10	****	0.070	****	****	****	0.108	****	0.030	****	****	****	0.062
25,15	****	0.098	****	****	0.165	0.144	****	0.029	****	****	0.080	0.068
50,15	****	****	****	****	****	0.150	****	0.056	****	****	****	0.071

Table 5. The power of rates of the all tests when $p=1$, $\beta_1 = 1$, $\beta_2 = 4$.

n	$\sigma = (1,4)$						$\sigma = (1,8)$					
	C	T	W	WF	PB	CAT	C	T	W	WF	PB	CAT
5,5	****	0.135	****	****	****	0.219	****	0.043	****	****	****	0.096
10,10	****	0.295	****	****	****	0.509	****	0.059	****	****	****	0.177
15,15	0.737	0.470	****	****	0.737	0.708	0.265	0.087	****	****	0.279	0.254
25,25	0.923	0.774	****	0.926	0.920	0.913	0.427	0.170	****	****	0.441	0.424
50,50	0.997	0.986	0.997	0.997	0.997	0.997	0.736	0.435	0.747	0.740	0.736	0.726
5,10	0.258	0.151	****	****	****	0.479	0.067	0.030	****	****	****	0.183
5,15	0.241	0.155	****	****	****	0.632	0.042	0.021	****	****	****	0.260
10,15	0.559	0.357	****	****	0.722	0.690	0.156	0.064	****	****	****	0.274
15,25	0.783	0.607	0.920	0.912	0.908	0.900	0.231	0.100	****	****	0.437	0.417
15,50	0.862	0.779	0.995	0.995	0.995	0.994	0.170	0.092	****	0.715	0.717	0.703
10,5	****	****	****	****	****	0.244	****	0.089	****	****	****	0.108
25,5	****	****	****	****	****	0.239	****	****	****	****	****	0.095
15,10	****	0.387	****	****	****	0.521	****	0.089	****	****	****	0.184
25,15	****	0.628	****	****	0.756	0.735	****	0.138	****	****	0.296	0.268
50,15	****	****	****	****	****	0.727	****	0.210	****	****	****	0.266

Table 6. The power of rates of the all tests when $p=2$, $\beta_1 = (1,1)$, $\beta_2 = (1,2)$.

n	$\sigma = (1,4)$						$\sigma = (1,8)$					
	C	T	W	WF	PB	CAT	C	T	W	WF	PB	CAT
5,5	****	0.043	****	****	****	0.055	****	0.031	****	****	****	0.048
10,10	****	0.032	****	****	****	0.071	****	0.019	****	****	****	0.061
15,15	****	0.033	****	****	****	0.092	****	0.009	****	****	****	0.063
25,25	****	0.039	****	****	****	0.151	****	0.011	****	****	0.084	0.074
50,50	0.307	0.098	****	****	0.314	0.292	0.117	0.013	0.131	0.121	0.117	0.107
5,10	0.023	0.010	****	****	****	0.082	0.017	0.007	****	****	****	0.058
5,15	0.009	0.004	****	****	****	0.099	0.003	0.002	****	****	****	0.063
10,15	0.056	0.021	****	****	****	0.099	0.030	0.005	****	****	****	0.063
15,25	0.054	0.015	****	0.169	0.162	0.140	0.024	0.006	****	****	0.085	0.071
15,50	0.016	0.007	****	0.279	0.278	0.259	0.003	0.001	****	0.121	0.119	0.110
10,5	****	****	****	****	****	0.056	****	****	****	****	****	0.046
25,5	****	****	****	****	****	0.057	****	****	****	****	****	0.049
15,10	****	0.051	****	****	****	0.070	****	0.027	****	****	****	0.059
25,15	****	0.063	****	****	0.127	0.100	****	0.028	****	****	****	0.059
50,15	****	****	****	****	****	0.099	****	0.044	****	****	0.074	0.056

Table 7. The power of rates of the all tests when $p=2$, $\beta_1 = (1,1)$, $\beta_2 = (1,4)$.

n	$\sigma = (1,4)$						$\sigma = (1,8)$					
	C	T	W	WF	PB	CAT	C	T	W	WF	PB	CAT
5,5	****	0.092	****	****	****	0.122	****	0.040	****	****	****	0.066
10,10	****	0.157	****	****	****	0.323	****	0.041	****	****	****	0.123
15,15	****	0.289	****	****	****	0.541	****	0.048	****	****	****	0.183
25,25	****	0.570	****	****	****	0.817	****	0.085	****	****	0.344	0.311
50,50	0.994	0.950	****	****	0.992	0.991	0.618	0.247	0.642	0.629	0.617	0.595
5,10	0.124	0.069	****	****	****	0.319	0.035	0.016	****	****	****	0.114
5,15	0.106	0.066	****	****	****	0.492	0.017	0.008	****	****	****	0.171
10,15	0.379	0.180	****	****	****	0.528	0.099	0.023	****	****	****	0.175
15,25	0.612	0.377	****	0.853	0.844	0.813	0.128	0.038	****	****	0.341	0.305
15,50	0.657	0.508	****	0.982	0.982	0.981	0.064	0.026	****	0.605	0.603	0.576
10,50	****	****	****	****	****	0.115	****	****	****	****	****	0.063
25,5	****	****	****	****	****	0.116	****	****	****	****	****	0.064
15,10	****	0.250	****	****	****	0.332	****	0.056	****	****	****	0.125
25,15	****	0.435	****	****	0.629	0.556	****	0.083	****	****	****	0.182
50,15	****	****	****	****	****	0.549	****	0.143	****	****	0.215	0.175

Table 8. The power of rates of the all tests when $p=3$, $\beta_1 = (1,1,1)$, $\beta_2 = (1,1,2)$.

n	$\sigma = (1,4)$						$\sigma = (1,8)$					
	C	T	W	WF	PB	CAT	C	T	W	WF	PB	CAT
10,10	****	0.030	****	****	****	0.052	****	0.015	****	****	****	0.048
15,15	****	0.026	****	****	****	0.075	****	0.012	****	****	0.076	0.058
25,25	****	0.027	****	****	0.156	0.121	****	0.006	****	****	****	0.066
50,50	0.254	0.056	****	0.278	0.267	0.232	0.104	0.011	****	0.108	0.099	0.088
10,15	0.044	0.013	****	****	****	0.081	0.030	0.006	****	****	****	0.058
15,25	0.038	0.008	****	0.158	0.150	0.119	0.018	0.003	****	****	0.082	0.065
15,50	0.007	0.003	****	0.240	0.237	0.217	0.001	0.000	****	0.099	0.096	0.085
15,10	****	0.062	****	****	****	0.062	****	0.029	****	****	****	0.053
25,15	****	0.059	****	****	****	0.084	****	0.021	****	****	****	0.058
50,15	****	****	****	****	****	0.105	****	0.045	****	****	****	0.057

Overall, the proposed CAT has type I error rates quite close to the nominal level α regardless of sample sizes, variance configurations, the number of explanatory variable and the relation of sample sizes and variances. Furthermore, the CAT appears to be more powerful than the other tests for all the considered situations.

Table 9. The power of rates of the all tests when $p=3$, $\beta_1 = (1, 1, 1)$, $\beta_2 = (1, 1, 4)$.

n	$\sigma = (\mathbf{1}, \mathbf{4})$						$\sigma = (\mathbf{1}, \mathbf{8})$					
	C	**T**	**W**	**WF**	**PB**	**CAT**	**C**	**T**	**W**	**WF**	**PB**	**CAT**
10,10	****	0.112	****	****	****	0.209	****	0.028	****	****	****	0.086
15,15	****	0.196	****	****	****	0.391	****	0.031	****	****	0.185	0.130
25,25	****	0.442	****	****	0.798	0.732	****	0.046	****	****	****	0.233
50,50	0.987	0.891	****	0.987	0.986	0.981	0.549	0.158	****	0.576	0.561	0.522
10,15	0.276	0.110	****	****	****	0.398	0.068	0.016	****	****	****	0.125
15,25	0.484	0.249	****	0.785	0.775	0.709	0.084	0.019	****	****	0.306	0.243
15,50	0.488	0.334	****	0.972	0.972	0.966	0.023	0.007	****	0.533	0.529	0.487
15,10	****	0.189	****	****	****	0.216	****	0.044	****	****	****	0.087
25,15	****	0.333	****	****	****	0.411	****	0.061	****	****	****	0.129
50,15	****	****	****	****	****	0.415	****	0.124	****	****	****	0.137

5. Numerical example

In this section of the paper, the use of test statistics given in Section 2 and the CAT statistic are demonstrated on numerical example. A real data about house insulation given by [15] was used. This data set was collected in the 1960s by Derek Whiteside of the UK Building Research Station. The weekly gas consumption (in 1000 cubic feet) and the average temperature (in degrees Celsius) were recorded for 26 weeks before and 30 weeks after the insulation of the cavity wall. This data set is given in Table 10.

Table 10. The dataset of average outside temperature and gas consumption for before and after insulation.

Before insulation	Average outside temperature $(0°\mathbf{C})$	-0.8, -0.7, 0.4, 2.5, 2.9, 3.2, 3.6, 3.9, 4.2, 4.3, 5.4, 6, 6, 6, 6.2, 6.3, 6.9, 7, 7.4, 7.5, 7.5, 7.6, 8, 8.5, 9.1, 10.2
	Gas consumption (1000cubic feet)	7.2, 6.9, 6.4, 6, 5.8, 5.8, 5.6, 4.7, 5.8, 5.2, 4.9, 4.9, 4.3, 4.4, 4.5, 4.6, 3.7, 3.9, 4.2, 4, 3.9, 3.5, 4, 3.6, 3.1, 2.6
After insulation	Average outside temperature $(0°\mathbf{C})$	-0.7, 0.8, 1, 1.4, 1.5, 1.6, 2.3, 2.5, 2.5, 3.1, 3.9, 4, 4, 4.2, 4.3, 4.6, 4.7, 4.9, 4.9, 4.9, 5, 5.3, 6.2, 7.1, 7.2, 7.5, 8, 8.7, 8.8, 9.7
	Gas consumption (1000cubic feet)	4.8, 4.6, 4.7, 4, 4.2, 4.2, 4.1, 4, 3.5, 3.2, 3.9, 3.5, 3.7, 3.5, 3.5, 3.7, 3.5, 3.4, 3.7, 4, 3.6, 3.7, 2.8, 3, 2.8, 2.6, 2.7, 2.8, 1.3, 1.5

We would like to find out how gas consumption is related with the outside temperature, that is, we aim to estimate a simple gas consumption function. Besides, it is also aimed to examine how this relationship changes after insulation. In other words, we tried to find out whether a structural change in the gas consumption occurred after the insulation was performed. To solve these problems, following functional relationship between the weekly gas consumption and the average outside temperature before (B) and after (A) wall insulation were specified, respectively.

$$(Gas\ consumption)_B = 6.854 - 0.393(average\ outside\ temperature)_B$$
$$(Gas\ consumption)_A = 4.724 - 0.278(average\ outside\ temperature)_A$$

The estimated sample variances are 0.079 and 0.126, respectively and the scatter plots of the data set is given in Figure 1. As seen in Figure 1, there is an evidence of homogeneity. In addition, the Levene test for homoscedasticity gives *p*-value as 0.545. A standard analysis yields a *p*-value of 0.000 for the common intercept hypothesis and 0.001 for the common slope hypothesis. That is, there is a significant difference in terms of the existence of insulation. The structural change in gas consumption regarding insulation was tested

by using the CAT, C, T, W, WF, and PB tests in Section 2. Table 11 presents the test statistics and *p*-values of the tests.
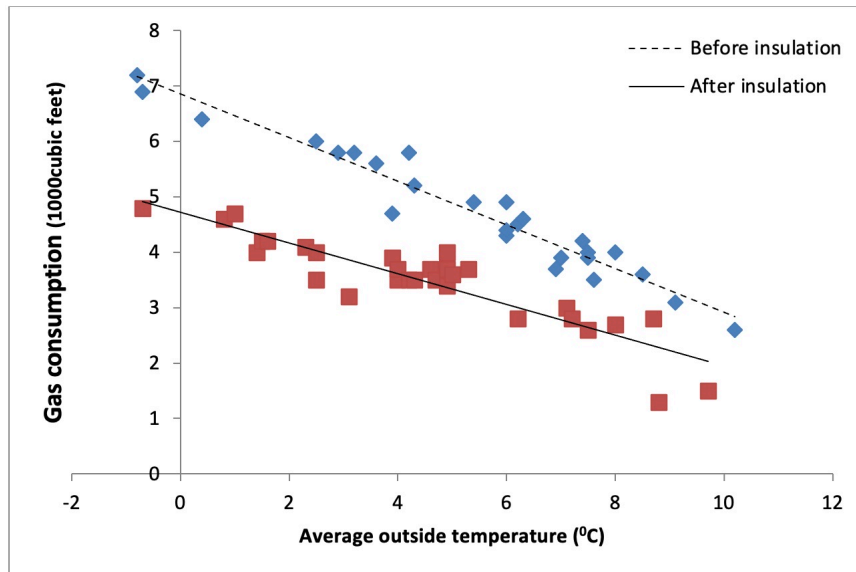

Figure 1. The scatter plots for data set given in Table 10.

Table 11. The results of all the tests.

	Chow	Toyoda	Wald	WF	PB	CAT
Test statistics	0.3278	0.3214	0.3528	0.3528	0.3528	0.3358
p-values	0.5693	0.5732	0.5525	0.5550	0.5548	0.6768

The *p*-values indicate that the tests do not reject the H_0 given in Eq. 2.2 at nominal level 0.05. That is, all tests indicate that there is no significant difference in the gas consumption function in terms of the existence of insulation. In this case, we may justify the use of pooled regression. Except for the CAT, *p*-values of all tests are close to each other. The CAT does not reject H_0 with greater p value compared to the other tests.

6. Discussion and conclusions

In this study, the problem of comparing two linear regression models when the variances are not assumed equal was considered. For this problem, a test based on the CAT method was proposed and was compared with the C, T, W, WF, and PB tests. In the simulation study, the explanatory variables $p=1$, $p=2$, and $p=3$ with different combinations of equal and unequal sample sizes and different variances were considered.

According to the simulation study, it is seen that the power and type I error rates of our proposed test is quite encouraging compared with the other tests reported in the literature. For this reason, the proposed CAT is a better alternative test for comparing two linear regression models under heteroscedasticity. Currently, we are also studying the test based on CAT for comparing several linear regression models under heteroscedasticity and hopefully this will be reported in a future publication.

The CAT method can be applied in many hypothesis testing problems with nuisance parameters. The success of the CAT depends on the selection of the test statistic and the resampling done from the restricted model for obtaining the bootstrap distribution of the test statistic, under null hypothesis. In CAT method, the RMLEs of parameters may not be available in closed form, for this reason, the iteration methods were frequently used.

However, it is difficult to apply the CAT method in some hypothesis testing problems in where RMLEs are not obtained even by iteration methods. For this reason, maybe it can be used an alternative statistic rather than the MLE and RMLE.

References

[1] M.M. Ali and J.L. Silver, *Tests for equality between sets of coefficients in two linear regressions under heteroscedasticity*, J. Amer. Statist. Assoc. **80** (391), 730-735, 1985.

[2] C.H. Chang, J.J. Lin and N. Pal, *Testing the equality of several gamma means: A parametric bootstrap method with applications*, Comput. Statist. **26** (1), 55-76, 2011.

[3] C.H. Chang and N. Pal, *A revisit to the Behren-Fisher Problem: Comparison of five test methods*, Comm. Statist. Simulation Comput. **37** (6), 1064-1085, 2008.

[4] C.H. Chang, N. Pal, W.K. Lim and J. J. Lin, *A note on comparing several Poisson means*, Comm. Statist. Simulation Comput. **39** (8), 1605-1627, 2010.

[5] G.C. Chow, *Tests of equality between sets of coefficients in two linear regressions*, Econometrica **28** (3), 591-605, 1960.

[6] M.D. Conerly and E.R. Mansfield, *An approximate test for comparing heteroscedastic regression models*, J. Amer. Statist. Assoc. **83** (403), 811-817, 1988.

[7] A.C. Davison and D.V. Hinkley, *Bootstrap Methods and their Application*, Cambridge University Press, 1997.

[8] T.J. Diciccio, M.A. Martin and S.E. Stern, *Simple and accurate one-sided inference from signed roots of likelihood ratios*, Canad. J. Statist. **29** (1), 67-76, 2001.

[9] E.Y. Gokpinar and F. Gokpinar, *A test based on the computational approach for equality of means under the unequal variance assumption*, Hacet. J. Math. Stat. **41** (4), 605-613, 2012.

[10] E.Y. Gokpinar, E. Polat, F. Gokpinar and S. Gunay, *A new computational approach for testing equality of inverse gaussian means under heterogeneity*, Hacet. J. Math. Stat. **42** (5), 581-590, 2013.

[11] F. Gokpinar and E. Gokpinar, *Testing the equality of several Log-normal means based on a computational approach*, Comm. Statist. Simulation Comput. **46** (3), 1998-2010, 2017.

[12] S.M. Goldfeld and R.E. Quandt, *Asymptotic tests for the constancy of regressions in the heteroscedastic case*, Econometric Research Program Research Memorandum, Princeton University **229**, 1978.

[13] S.A. Gupta, *Testing the equality between sets of coefficients in two linear regressions when disturbances are unequal*, PhD thesis, Purdue University, 1978.

[14] H. Haghbin, M.R. Mahmoudi and Z. Shishebor, *Large sample inference on the ratio of two independent binomial proportions*, J. Math. Ext. **5** (1), 87-95, 2015.

[15] D.J. Hand, F. Daly, A.D. Lunn, K.J. McConway and E. Ostrowski, *A Handbook of Small Data Sets*, Chapman and Hall, 1994.

[16] Y. Honda, *On tests of equality between sets of coefficients in two linear regressions when disturbance variances are unequal*, The Manchester School **50** (2), 116-125, 1982.

[17] W.A. Jayatissa, *Tests of equality between sets of coefficients in two linear regressions when disturbance variances are unequal*, Econometrica **45** (5), 1291-1292, 1977.

[18] K.R. Kadiyala and S. Gupta, *Tests for pooling cross-sectional data in the presence of heteroscedasticity*, Bulletin of the Institute of Mathematical Statistics **7**, 275-276, 1978.

[19] K. Knight, *Mathematical Statistics*, Chapman and Hall, 1999.

[20] M.R. Mahmoudi, J. Behboodian, and M. Maleki, *Large sample inference about the ratio of means in two independent populations*, J. Stat. Theory Appl. **16** (3), 366-374, 2017.

[21] M.R. Mahmoudi and M. Mahmoodi, *Inferrence on the ratio of variances of two independent populations*, J. Math. Ext. **7** (2), 83-91, 2014.

[22] M.R. Mahmoudi, A.R. Nematollahi and A.R. Soltani, *On the detection and estimation of the simple harmonizable processes*, Iran. J. Sci. Technol. Trans. A: Sci. **39** (2), 239-242, 2015.

[23] E. Moreno, F. Torres and G. Casella, *Testing equality of regression coefficients in heteroscedastic normal regression models*, J. Statist. Plann. Inference **131** (1), 117-134, 2005.

[24] D. Oberhelman and R. Kadiyala, *A test for the equality of parameters for separate regression models in the presence of heteroscedasticity*, Comm. Statist. Simulation Comput. **36** (1), 99-121, 2007.

[25] K. Ohtani and T. Toyoda, *A Monte Carlo study of the Wald, LM, and LR tests in a heteroscedastic linear model*, Comm. Statist. Simulation Comput. **14** (3), 735-746, 1985.

[26] N. Pal, W.K. Lim and C.H. Ling, *A computational approach to statistical inferences*, J. Appl. Probab. **2** (1), 13-35, 2007.

[27] J.J. Pan, M.R. Mahmoudi, D. Baleanu and M. Maleki, *On comparing and classifying several independent linear and non-linear regression models with symmetric errors*, Symmetry **11** (6), 820, 2019.

[28] S.M. Sadooghi-Alvandi, A.A. Jafari, and H.A. Mardani-Fard, *Comparing several regression models with unequal variances*, Comm. Statist. Simulation Comput. **45** (9), 3190-3216, 2016.

[29] P. Schmidt and R. Sickles, *Some further evidence on the use of the Chow test under heteroscedasticity*, Econometrica **45** (5), 1293-1298, 1977.

[30] A. Schworer and D.P. Hovey, *NewtonRaphson versus fisher scoring algorithms in calculating maximum likelihood estimates*, in: Electronic Proceedings of Undergraduate Mathematics Day, University of Dayton, **1**, 111, 2004.

[31] J.G. Thursby, *A comparison of several exact and approximate tests for structural shift under heteroscedasticity*, J. Econometrics **53** (1-3), 363-386, 1992.

[32] JL. Tian, C. Ma and A. Vexler, *A parametric bootstrap test for comparing heteroscedastic regression models*, Comm. Statist. Simulation Comput. **38** (5), 1026-1036, 2009.

[33] T. Toyoda, *Use of the Chow test under heteroscedasticity*, Econometrica **42** (3), 601608, 1974.

[34] P.A. Watt, *Tests of equality between sets of coefficients in two linear regressions when disturbance variances are unequal: Some small properties*, The Manchester School **47** (4), 391-396, 1979.

[35] J. Xu and X. Li, *A fiducial p-value approach for comparing heteroscedastic regression models*, Comm. Statist. Simulation Comput. **47** (2), 420-431, 2018.