Performance of Methods Determining Structural Break in Linear Regression Models

Zümre Özdemir Güler[®] and Mehmet Akif Bakır

Karamanoglu Mehmet Bey University and Gazi University

ABSTRACT

In the literature, although many studies are describing the structural break in the linear regression model with time-series data, studies investigating this issue with cross-sectional data are limited. In this study, the performance evaluation of some approaches used to determine the structural break in a linear regression equation based on cross-sectional data was performed. In this context, firstly, the structural break problem is defined. Then, the theoretical expositions of some well-known methods which determine the structural break are given. The methods which used to determine the structural breaks may show performance differences under the effect of some factors. The performances of selected methods were evaluated with a simulation study in the context of the difference of constant terms, the difference of slope coefficients, location of break-point, sample size and homogeneity of error variances. The results of the simulation showed that the performances become different in terms of some structural features from the suggested methods for determination of structural break.

Key words: *Structural Break, Linear Regression Models, Cross-Sectional Data, Break-Point*
JEL Classifications: C40, C530

1. INTRODUCTION

In the literature, it is known that many methods determine the structural break in the regression model. The majority of these methods are based on the use of time-series data. However, the number of studies based on cross-sectional data is limited. In this context, this study focuses on comparing the performances of the methods which determine the break in the linear regression model using cross-sectional data. A structural break in the regression equation can be defined as the change of the relationship form between dependent and independent variables for certain values of independent variables. In the case of time-series data, the structural break is usually defined as a change from one period to another. Therefore, the model parameters do not remain same throughout the sample period. This can be expressed as "structural instability" in the regression equation. This instability means that there are two or more different behavioral patterns during the sample period. In the literature, the problem of structural break is also expressed in terms such as "stability of regression equation", "parameter stability", "structural change", "structural determination", "structural change".

[®] Zümre Özdemir-Güler, Research Asistant of Econometrics, Karamanoğlu Mehmet Bey University, Karaman, TURKEY, 70200, (email: zumreozd@kmu.edu.tr), Tel: +903382262000, Fax: +903382262023.
Mehmet Akif Bakır, Professor of Statistics, Gazi University, Ankara, TURKEY, 06500, (email: mabakir@gazi.edu.tr), Tel: +903122021470, Fax: +903122122279.

The structural break problem is generally seen in time-series data, but it can also occur in the regression equations with the cross-sectional data. For example, in terms of time-series data, from a macroeconomic context, a society's saving trends may differ before and after the period of crisis. On the other hand, in the context of cross-sectional data, the saving trends of the socio-economic groups may be different. Saving trends of social socio-economic sections may be different from each other in cross-section data. Structural break problem is one of the most important discussion areas in terms of both time series and regression analysis. However, in this study, only the problem of a structural break in the regression equation is considered and the problem of a break in time series model is ignored.

In the regression analysis, it is generally assumed that the dependent variable is a function of the independent variable during the sample period, and the behavioral model does not change during this period. In the same sample period, there may be structural differences between the dependent variable and the independent variables. In this case, estimating a single regression equation for the sample period may reduce the reliability of parameter estimates. It can also reduce the forecast and prediction performance of the estimation equation. Fitting a single regression equation for a sample period in the case of structural break may cause problems about estimation. Also, whether there is a structural break in the data set is an important problem especially in economic studies.

In a structural break, the sample can be separated into two different data sets (one break) as well as it can be included in more than two sets (multiple breaks). "Multiple breaks" can be discussed especially when the examined sample is big enough. However, this study was limited only the case that has a single break.

Many studies determine both single and multiple breaks for time series models. However, there are relatively limited number of approaches to a single break for cross-sectional data regression models. Nevertheless, the recommended methods may not work effectively in all situations. Therefore, the tests developed based on specific criteria have weaknesses or strengths. Under the same conditions, the superiority of a test against others is measured by having the maximum power (Hansen, 1992).

The structural break tests used in the literature can be classified into two groups. The methods in the first group determine the existence of structural break and the second group determine its location in addition to the presence of break. The cases which the structural break is unknown, using the test from the second group becomes important.

Approaches to the determination of structural breaks work under certain assumptions. A structural break in a regression equation occurs when there are significant differences in terms of constant coefficients, slope coefficients, or both. Sample size and sub-sample size may have an impact on the performances of these methods which work based on the separation of the data set into subsets. The homogeneity in the error variance of the subsets may be effective on the model behaviors, too. Therefore, the performance of these methods may be considered as a function of five main factors. These are the size of the absolute difference between constant terms, the size of the absolute difference between slope coefficients, the homogeneity of error variances, the sample size and the location of the break-point. In this context, the performance evaluation of the methods in terms of the power of the test was carried out with a full factorial experiment design based on the five factors listed.

In the study, Chow Break-point Test (*CBT*), Chow Predictive Test (*CPT*), Chow Test with Heteroscedasticity (*HCT*), Chow Test with Dummy Variables (*DCT*), Hendry Predictive Test (*HPT*), Recursive Residual Test (*RRT*), CUSUM Test (*CST*), CUSUM^2 Test (*CSQT*) and *MZ* Test which are well known in the literature and used in many software, are compared. Performance evaluation was made on the synthetic data sets which were randomly produced.

The rest of this paper is structured as follows. Firstly, the concept of structural break is defined and methods for determining the structural break are clarified. Then, the simulation study carried out taking into account the model combinations formed by the experimental design method is explained. After that, the effects of the main factors and the binary interactions of the factors on the structural break methods are revealed. Finally, the findings of the analyzes were discussed.

2. STRUCTURAL BREAK AND METHODS OF DETERMINING STRUCTURAL BREAK

In multiple regression models, it is assumed that the model parameters will not change over the sample. However, economic behaviors may change in time due to different reasons such as changes in economic policy, changes in economic structure, technological developments, economic crisis, and postwar reconstruction. Besides, the consumption patterns of households may be shown by different behavior structures as to income groups. Due to these reasons, there may be significant differences in the distribution of the dependent variable (especially in its mean). These breaks separate the data set into different parts having different characteristics. Therefore, the model's behavior in the sample period cannot be described with a single model. Different model characters make the description of the model more difficult.

In other words, the structural break will cause a break on the regression line in the linear regression model. This break can result in simultaneous differentiation of the constant terms or slope coefficients of the sub-sample periods or both of them. If the model is estimated without considering the break, the model's functional shape will be faulty and a description error will occur. Thus, the model must be estimated in at least two parts (or more).

Various approaches have been developed to determine structural break in linear regression models. Each model has different requirements and capabilities. Most widely known among these is the Chow Test (*CT*) which suggested by Gregory Chow (1960). This test aims to determine one break or multiple breaks in the regression equation. *CT* tests the null hypothesis of "no structural break" against the alternative hypothesis of "there is a structural break". For the *CT* to be valid some assumptions must be provided. These assumptions can be summarized as follows for the model in which there is one break:

- Error terms of the sub-samples (ε_{t1} and ε_{t2}) must be independent from each other.

- Each error term of sub-samples should be distributed normally with zero mean and constant variance,
$$\varepsilon_{t1} \sim N(0, \sigma^2) \text{ and } \varepsilon_{t2} \sim N(0, \sigma^2)$$

- Break-point must be known in advance.

The fact that the test is based on assumptions in this way may reveal some weaknesses. For example, this cannot apply when assumptions about error variances and the normality of error

terms are not provided or the breakpoint is not known in before. Furthermore, the test does not give any information about where the difference in the regression model is originated and this is one of its weaknesses. On the other hand, it can be said that *CT* is more useful than other tests due to its ease of operation and feasibility in case of more than one break-point.

Chow, suggests two methods: *CBT* and *CPT*. Both of these tests use *F* statistic which is identified as based on the Sum of Squared Residual (*SSR*). The proposed tests divide the data set into two from the known break-point to determine the structural break. Then, they compare their *SSR*s.

While testing for structural breaks with *CBT*, the data set containing *n* observations are divided into two sub-sample sets containing n_1 and n_2 observations. If it is considered that the model estimates *k* number of parameters, for estimation of the sub regression equations, (n_1-k) and (n_2-k) must be greater than 0. Consequently, it must be $n_1 > k$ and $n_2 > k$. A deficiency of *CBT* is the requirement that each subsample's observation number must be at least the same with the number of parameters. Because, when there is a structural break and one of the subsets have fewer observations, these conditions may not be provided and it may be $n_1 < k$ and $n_2 < k$. In such circumstances, instead of *CBT*, *CPT* must be used.

CPT is a test that shows whether the results obtained from one of the subsamples are valid for the whole sample set. The idea behind this recommendation is to compare the estimated *SSR* with the *SSR* obtained from one of the sub-samples (larger than the number of observations) and for all sample observations.

CT reflects well behavior even under heteroscedasticity when at least one of the two subsamples is bigger. For two sample orders with small sizes, the test's significance level will be affected more in case that heteroscedasticity is more moderate. The level of the real significance under heteroscedasticity is always bigger than so-called the significance level of homoscedasticity deemed as a prerequisite (Toyoda, 1974).

As in econometric research, when *CT* is applied for smaller samples, it may be useful to make some prior tests against heteroscedasticity of error terms. If these tests show a presence of heteroscedasticity, *CT* must be used directly. In this case, the test can be applied after providing homoscedasticity of the error terms obtained by multiplying the original regression equation with an appropriate matrix (Toyoda, 1974).

Error variances must be homogenous in all of these approaches. Toyoda (1974) has shown that *CT* is unsuccessful when these assumptions are not provided and the sample size is smaller.

All of the methods mentioned above determine whether there is a break in the regression equation but they do not give information on constant term or slope coefficient from which the behavior change in equation come. The *DCT* method, which requires the prior knowledge of the break-point, can be used to overcome this deficiency. In this method, the dummy variable which is equal to the subsamples can be used. The significance of dummy and interaction variable coefficients indicates the presence and source of the structural break.

Both *CT* and *DCT* give the same results. However, *DCT* has more advantages than *CT*. Firstly, despite the *CT* can determine the difference between the two regression lines, it cannot determine where this difference originates. Both the presence of the structural break and the

cause of the break can be determined using the *DCT*. Another advantage of *DCT* is the ease of application. All the information needed can be obtained from a single regression equation with *DCT* while requiring a multi-stage process at *CT*. Moreover, the parameters of the two sample sets can be estimated from a regression equation (Gujarati, 1970).

Similar to the Chow test, Hendry et al. (1989) also proposed another structural break method that requires prior knowledge for the break-point, and this method is called Hendry Prediction Test (*HPT*). This method tests the null hypothesis of homoscedasticity against the alternative hypothesis of heteroscedasticity. Test statistics has χ^2 distribution. Furthermore, Maasoumi et al. (2010) developed *the MZ* Test for structural change with a known break-point when not only the regression coefficients but also the error variances change. The most important feature of the *MZ* test is that it tests the null hypothesis that there is no structural break versus the alternative hypothesis that break in both regression coefficients and the error variances.

As with most of the economic problems, prior information on the structural break-point is known by the researcher. However, some stochastic processes, such as quality control processes, may not need prior knowledge of the structural breakpoint. In such cases, the location of the break-point is considered as a distinct problem. When prior information on the break-point is not present, the approaches mentioned above cannot be used. For the removal of this problem, three different methods were developed by Brown et al. (1975): *RRT, CST,* and *CSQT*. *RRT* is based on an addition of data to the small sample and "recursive residuals" (*rr*) which were obtained by using the *OLS* estimations on the previous point. *CST* is based on standardized *rr* obtained by standardizing *rr* obtained in *RRT*. Finally, *CSQT* determines the structural break and its place by using standardized *rr*'s squares.

Before making performance evaluation works of the aforementioned methods, it would be useful to make a summary of them. In this section of the study, theoretical and functional frames of the methods which were suggested and performance-evaluated for the determination of the behavioral differences in the model will be discussed.

2.1. Chow Test

The principle of the method is based on testing null hypothesis "H_0: There is no structural break." against alternative hypothesis "H_1: There is a structural break". The method separates the considered data set into subsets and these subsets are based on the comparison of *SSR*s obtained by estimation. Consider the following linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (i=1,2,\dots,n) \tag{2.1}$$

When the data are separated from break-point into two independent data sets which consist the number of n_1 and n_2 observations ($n=n_1+n_2$), regression equations for these two data sets can be identified as

$$Y_{i1} = \beta_{01} + \beta_{11} X_{i1} + \varepsilon_{i1} \quad (i=1,2,\dots, n_1) \tag{2.2}$$
$$Y_{i2} = \beta_{02} + \beta_{12} X_{i2} + \varepsilon_{i2} \quad (i=1,2,\dots, n_2) \tag{2.3}$$

The procedure applied based on the size of subsamples can be classified as *CBT* and *CPT*. *CBT* is a method based on the principle of the proportion of *SSR*s which obtained from these two equations with each other. *CBT* statistic is calculated as:

$$F_{CBT} = \frac{[\mathbf{e}'_{UR}\mathbf{e}_{UR} - (\mathbf{e}'_1\mathbf{e}_1 + \mathbf{e}'_2\mathbf{e}_2)]/k}{(\mathbf{e}'_1\mathbf{e}_1 + \mathbf{e}'_2\mathbf{e}_2)/(n-2k)} \tag{2.4}$$

where $\mathbf{e}_{UR}$ vector is $n \times 1$ vector which consists of the estimated residual values of the

unconstrained model for all sample observations. $\mathbf{e}_1$ is $n_1 \times 1$ vector which consists of the estimated residual values of the constraint model for the first subsample. $\mathbf{e}_2$ is $n_2 \times 1$ vector which involves the estimated residual values obtained from the constraint model for the second sub-sample.

The degree of freedom for the first sub-sample (df_1) is $n_1 - k$, the degree of freedom for the second subsample (df_2) is $n_2 - k$ and the denominator's degree of freedom is $n_1 - k + n_2 - k = n - 2k$. The numerator's degree of freedom is k, which is the number of parameters in the model. For this reason, the distribution of the test statistic (F_{CBT}) is F distribution with degrees of freedom k and $n - 2k$ ($F_{CBT} \sim F_{\alpha;k,n-2k}$). If $F_{CBT} > F_{\alpha;k,n-2k}$, the null hypothesis H_0 is rejected.

CPT estimates two models. First of these models is the model using all data and the second one is the model using the bigger sub-sample data. If it is supposed that the bigger sub-sample is the first sub-sample (n_1), first the regression equation is estimated by using n_1 observation. This estimated equation is used for predicting n_2 observation. Then the hypothesis that the means of error terms are "zero" is tested. In this case, the *CPT* statistic is:

$$F_{CPT} = \frac{(\mathbf{e}'_{UR}\mathbf{e}_{UR} - \mathbf{e}'_1\mathbf{e}_1)/n_2}{\mathbf{e}'_1\mathbf{e}_1/(n_1-k)} \qquad (2.5)$$

and F_{CPT} statistic has F distribution with degrees of freedom n_2 and $n_1 - k$ ($F_{CPT} \sim F_{\alpha;n_2,n_1-k}$). The calculated test statistic will be compared to the following with type I error α, $df_1 = n_2$ and $df_2 = n_1 - k$, the regression equation is estimated by using n_1 observation. If $F_{CPT} > F_{\alpha;n_2,n_1-k}$ the null hypothesis is rejected and it is interpreted as "there is a structural break in the regression model".

2.2. Chow Test in Case of Heteroscedasticity

In the case of heteroscedasticity in the model, Generalized Least Square (*GLS*) method will be used instead of the *OLS* method. *GLS* is the estimation of the regression equation after the variables of the model have been converted to provide the homoscedasticity assumption of error variances. The conversion of variables is made by dividing each variable in the model by error variance (standard deviation) $w_i = 1/\sigma_i$ or by another suitable weighting (Gujarati, 2004).

$$\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\beta} + \mathbf{e}^* \qquad (2.6)$$

where $\mathbf{Y}^* = \frac{1}{\sigma}(\mathbf{Y}_{n\times1})$, $\mathbf{X}^* = \frac{1}{\sigma}(\mathbf{X}_{n\times k})$, $\boldsymbol{\beta}$ is $k \times 1$ coefficients vector for all samples and $\mathbf{e}^* = \frac{1}{\sigma}(\mathbf{e}_{n\times1})$.

$$\mathbf{Y}^*_t = \mathbf{X}^*_t\boldsymbol{\beta} + \mathbf{e}^*_t \quad (t=1,2) \qquad (2.7)$$

where $\mathbf{Y}^*_t = \frac{1}{\sigma_t}(\mathbf{Y}_{n_t\times1})$, $\mathbf{X}^*_t = \frac{1}{\sigma_t}(\mathbf{X}_{n_t\times k})$, $\boldsymbol{\beta}_t$ is $k \times 1$ coefficients vector for all sub-samples and $\mathbf{e}^*_t = \frac{1}{\sigma_t}(\mathbf{e}_{n_t\times1})$.

When the residuals of the estimated model for the whole sample are denoted by $\mathbf{e}^*_{UR}$ and the residuals of the models estimated for the sub-samples are denoted by $\mathbf{e}^*_1$ and $\mathbf{e}^*_2$, respectively, *HCT* statistic is calculated as follows:

$$F^*_{HCT} = \frac{[\mathbf{e}^{*'}_R\mathbf{e}^*_R - (\mathbf{e}^{*'}_1\mathbf{e}^*_1 + \mathbf{e}^{*'}_2\mathbf{e}^*_2)]/k}{(\mathbf{e}^{*'}_1\mathbf{e}^*_1 + \mathbf{e}^{*'}_2\mathbf{e}^*_2)/(n-2k)} \qquad (2.8)$$

If $F^*_{HCT} > F_{\alpha;k,n-2k}$ then it is interpreted as "there is a structural break in the regression model".

2.3. Dummy Variable Approach in case of Structural Break

In the case of a structural break, firstly, the regression equation with the dummy variable(s) is estimated for the whole data set. Then the significance of the coefficient of the dummy variable and the significance of the interaction coefficient is tested. The presence of a structural break and where it originated is determined by this test. Dummy variable added regression model is as follows:

$$Y_i = \beta_{01} + \beta_{02}D_i + \beta_{11}X_i + \beta_{12}(D_iX_i) + \varepsilon_i \quad (i=1,2,\dots,n) \tag{2.9}$$

where, D_i is a dummy variable and it takes the value of 0 for the period before the structural break, while it takes the value of 1 for the period after the structural break (Gujarati, 1970). Under the assumption $E(\varepsilon_i)=0$, when equation (2.9) is estimated, the following functions are obtained:

$$D_i=0 \Rightarrow E(Y_i|X_i) = \beta_{01} + \beta_{11}X_i \tag{2.10}$$

$$D_i=1 \Rightarrow E(Y_i|X_i) = (\beta_{01} + \beta_{02}) + (\beta_{11} + \beta_{12})X_i \tag{2.11}$$

These functions are the conditional average functions of Y given X and they provide the comparisons of sub-samples in which dummy variable gets values 0 (before the structural break) and 1 (after the structural break, Gujarati, 2004).

In equation (2.9), β_{02} shows the difference of the constant terms and β_{12} shows the difference of the slope coefficients. *DCT* is considered as two different tests. One of these tests determines a structural break which is caused by the difference of the constant terms (*DCT*–1) and the other one determines a structural break which is caused by the difference of slope coefficients (*DCT*–2).

2.4. Hendry Prediction Test

In *HPT*, firstly the equation (2.2) having the regression model of the first sub-sample is estimated. Then, the regression model of the second sub-sample and the coefficient estimations of the first sample b_{01} and b_{11} are used. The one-step prediction errors can be calculated as follows:

$$\mathbf{e}_2 = \mathbf{Y}_2 - b_{01} - b_{11}\mathbf{X}_2 \tag{2.12}$$

these one-step prediction errors can be used for calculation of *HPT* statistics. *HPT* statistics can be calculated as:

$$HPT = \frac{\mathbf{e}_2'\mathbf{e}_2}{\mathbf{e}_1'\mathbf{e}_1 / (n_1 - k)} \tag{2.13}$$

This calculated test statistic has Chi-square distribution with a degree of freedom n_2 $\left(HPT \sim \chi^2_{\alpha,n_2}\right)$.

2.5. Recursive Residual, *CUSUM* and *CUSUM²* Tests

For *RRT*, the regression set is estimated through *OLS* by using a small n_1 sized observation set. The observation number to be taken at the beginning (n_1) must be at least as much as the number of parameters k ($n_1 \geq k$). Then *OLS* estimations are repeated by increasing the data period one by one. Thus, $n-k+1$ coefficient estimation set is obtained.

Under the accuracy of the null hypothesis H_0, $\mathbf{b}_r$ is an *OLS* estimator of β based on r observations, the following formula is obtained:

$$\mathbf{b}_r=(\mathbf{X}_r'\mathbf{X}_r)^{-1}\mathbf{X}_r'\mathbf{Y}_r \tag{2.14}$$

where $\mathbf{X}_r'=[x_1,\dots,x_r]$ and $\mathbf{Y}_r'=[y_1,\dots,y_r]$, and $(\mathbf{X}_r'\mathbf{X}_r)$ is a non-singular matrix. In each step, the last estimation of $\mathbf{b}_r$ is used for the estimation of the next value of the dependent variable. The one-step prediction error obtained from this prediction is called recursive residual (rr). Thus, r^{th} term error prediction is calculated as follows:

$$\mathbf{v}_r=\mathbf{y}_r-\mathbf{x}_r'\mathbf{b}_{r-1} \quad (r=k+1,\dots,n) \tag{2.15}$$

Instead of rr ($\mathbf{v}_r$) in equation (2.15), standardized rr ($\mathbf{w}_r$) is calculated as:

$$\mathbf{w}_r=\frac{\mathbf{v}_r}{\sqrt{Var(v_r)}/\sigma}=\frac{\mathbf{y}_r-\mathbf{x}_r'\mathbf{b}_{r-1}}{\sqrt{\left(1+\mathbf{x}_r'\left(\mathbf{X}_{r-1}'\mathbf{X}_{r-1}\right)^{-1}\mathbf{x}_r\right)}} \quad (r=k+1,\dots,n) \tag{2.16}$$

$\mathbf{w}_r$ is standardized error estimator of $\mathbf{y}_r$ which was estimated from $y_1,\dots,y_{r-1}$. In equation (2.16) the variance of rr, $Var(v_r)$, is defined as:

$$Var(v_r)=\sigma^2(1+\mathbf{x}_i(\mathbf{X}_{i-1}'\mathbf{X}_{i-1})^{-1}\mathbf{x}_i') \tag{2.17}$$

where σ^2 means the error variance in regression period. $\mathbf{x}_i\,\mathbf{x}_i$ shows i^{th} term row vector of $\mathbf{X}$ matrix.

When rr (v_r) is used, the graph of v_r is plotted against time. In this graph, the lower and upper limits based on the following formula are drawn.

$$E(v_r)\pm2\sqrt{Var(v_r)} \tag{2.18}$$

If v_r's are exceeded these limits, it is decided that there is a structural break in the model. This is also equivalent to applying a t-test for each point.

When standardized rr (w_r) is used, these limits are considered as $+2$ and -2, because of $w_r\sim N(0,1)$. If w_r's are out of these limits, then it is concluded that there is a structural break.

In *RRT*, the moments when standardized residuals or calculated residuals get beyond their own limits, they are determined as the break-point and it is said that there is a structural break after this point.

In this part, for not creating confusions in the indexes, the standardized residuals will be shown with w_i instead of w_r. In *CST*, w_r has a normal distribution $N(0,\sigma^2)$ under H_0. w_r and w_i ($r\neq i$) are also independent. In this case, the values of *CUSUM* are calculated as follows:

$$CUSUM_r=\sum_{i=k+1}^{r}\frac{w_i}{\hat\sigma} \quad (r=k+1,\dots,n) \tag{2.19}$$

where

$$\hat\sigma^2=\frac{1}{n\text{-}k\text{-}1}\sum_{r=k+1}^{n}(w_r-\overline{w})^2 \tag{2.20}$$

and

$$\overline{w}=\frac{\sum_{i=1}^{n}w_i}{n\text{-}k} \tag{2.21}$$

lines relating $\pm a\sqrt{n-k}$ determinated at k point and $\pm3a\sqrt{n-k}$ determinated at n point are obtained in the graph drawn (Guris et al., 2010).

Parameter a is determined by the selected significance level. According to different significance levels, the value was given in Brown et al (1975) and a values which are frequently used for significance levels were given below:

$$\alpha=0.01 \quad \Rightarrow \quad a=1.143$$
$$\alpha=0.05 \quad \Rightarrow \quad a=0.948$$
$$\alpha=0.10 \quad \Rightarrow \quad a=0.850$$

If CUSUM values exceed any of lower and upper limits in the graph, the hypothesis on "*no structural break*" is rejected (Guris et al., 2010).

The hypothesis in *CSQT*, which is an alternative test similar to *CST*, are determined as in *CST* but test statistics are calculated by using standardized squares of *rr* in this test. For *CSQT*,

$$CUSUM_r^2 = \frac{\sum_{i=k+1}^{r} w_i^2}{\sum_{r=k+1}^{n} w_r^2} \quad (r=k+1,\ldots,n) \tag{2.22}$$

The values are calculated and the graph of $CUSUM_r^2$ is drawn. The expected value of the test statistic,

$$E(CUSUM_r^2) \cong \frac{r-k}{n-k} \tag{2.23}$$

When $r=k$ the expected value is "0" and $r=n$ it becomes "1". Confidence limits are determined as in *CST*. In *CSQT*, the confidence limits are obtained from the following equation:

$$E(CUSUM^2) \pm c_0 \tag{2.24}$$

Tables of c_0 values were given in Brown et al. (1975). c_0 is determined according to *m* and *α* significance level values according to whether the test has single or double sides. If $n-k$ value is an odd number, *m* value is calculated as:

$$m = \frac{1}{2}(n-k) - 1 \tag{2.25}$$

If $n-k$ value is an even number, *m* value is calculated and based interpolation is carried out as in the following (Guris et al., 2010).

$$m = \frac{1}{2}(n-k) - \frac{3}{2} \text{ and } m = \frac{1}{2}(n-k) - \frac{1}{2} \tag{2.26}$$

If the test is double-sided, the value of the c_0 is obtained with *m* and $\alpha/2$ values, if the test single-sided the c_0 value is calculated from *m* and *α* values. The upper and lower confidence limits of *CSQT* are drawn with the values determined from the table (Brown et al., 1975). If *CSQT* statistics cut the confidence limits, it is decided that there is a structural break.

2.6. *MZ Test*

In *MZ* Test, the standard linear regression model in equation (2.1) and separated models before and after break-point in equation (2.2) and (2.3) are considered. Under the null hypothesis of no structural break the usual *OLS* estimates of the parameters β and σ^2 are

$$\hat{\beta}_0 = (X'X)^{-1}X'Y \text{ and } \hat{\sigma}_0^2 = Y - X\hat{\beta}_0^2/(n-k),$$

respectively. And the standard *OLS* estimates for two models (eq. 2 and 3) are

$$\hat{\beta}_i = (X_i'X_i)^{-1}X_i'Y_i \ (i=1,2) \text{ and } \hat{\sigma}_i^2 = Y_i - X_i\hat{\beta}_i^2/(n_i-k) \ (i=1,2),$$

respectively (Ahmed et al., 2017).

MZ Test is calculated that a test of the null hypothesis "$H_0: \beta_1 = \beta_2, \sigma_1^2 = \sigma_2^2$" against alternative hypothesis "$H_1: \beta_1 \neq \beta_2$ or $\sigma_1^2 \neq \sigma_2^2$".

The *MZ* statistics defined as follows:

$$MZ = (n-k)\log\hat{\sigma}_0^2 - \{(n_1-k)\log\hat{\sigma}_1^2 + (n_2-k)\log\hat{\sigma}_2^2\} \tag{2.27}$$

The critical values for *MZ* Test are computed under the null hypothesis there is no structural break. To calculate the critical values we consider the model in equation (2.1). We generate independent variable *X* as a series ranging from 1 to 50 with a step size of 0.5 ($n=100$). We set $\beta_0=\beta_1=1$ and generate errors (ε) from the standard normal distribution. The dependent variable *Y* is obtained from equation (2.1). This simulated data is used to generate critical values which are got by performing 30.000 replications. The *MZ* Test rejects the null hypothesis when *MZ* statistics exceed the simulated critical value at 5% significance level.

3. EXPERIMENT DESIGN AND SIMULATION STUDY

In this part of the study, simulation studies were made under various conditions to make the performance evaluation for the structural tests considered in the previous section. The tests suggested for the determination of the structural break may have performance differences under some factors. These factors, as mentioned by Antoshin et al. (2008), can be listed as the size of the difference between constant terms and slope coefficients in the sub-sample regression equation, the division rate of data set in the break-point, sample observation coefficient and variables of sub-sample error variances. The methods discussed in this study will be compared with their performances which are represented as the power of test with regards to these factors.

The data used in this paper were produced by considering a single variable linear regression model. For determining that, there is a structural break in the model, an *n* sized sample was produced as a total of two different samples sized n_1 and n_2. When the data were produced, the assumptions of the linear regression model were considered. For instance, error terms were randomly produced from the normal distribution. If the explanatory variables are (X_i) average 0 and its variance 1 were produced randomly following the Standard Normal Distribution. Constant term (β_0) and slope coefficient (β_1) were produced in a manner to differentiate between 1st and 2nd samples. Then these obtained values are located in the models to determine the variable (*Y*).

Table 3.1 shows the factors included in the experimental design and their levels. In this study, the structural break in the model was created by taking into account 5 factors such as differentiation of constant term (*dc*), differentiation of slope coefficient (*ds*), location of break-point (*bp*), sample size (*ss*) and homogeneity of error term variance (*de*). In this context, model combinations consist from five-factors were studied by experimental design method. The differentiation in constant terms and slope coefficients was calculated as the absolute difference of the coefficients of the two sub-samples and it was discussed at three different levels as 0.00, 0.50 and 1.00. Here, the first level (0.00) means that there is no difference between the coefficients, i.e. there is no structural break in the model. The location of the break-point was calculated by the ratio of the sample size of the first sub-sample to the whole sample size and it was examined at three different levels, 0.10, 0.25 and 0.50. Thus, the question that "is the effect of the break of the data set in the early, midway or late?" was examined. The sample size was questioned at two levels, 60 and 120. Thus, the effect of small sample or large sample on the tests could be determined. The homogeneity of the error terms variance was discussed at two levels as 0.00 and 0.30. Here, the first level (0.00) refers to homoskedasticity and the second level (0.30) refers to heteroskedasticity. Thus, the effect of the homogeneity of error variance on the tests was observed. In this case, the total experiment number of the simulation study is $3^3 \times 2^2 = 108$ (three factors are consisting of three levels and two factors are consisting of two levels). The simulation study was carried out for a 5-factor

balanced full-factorial experiment. For each cell of the design matrix, the simulation operation was repeated 3000 times and the averages of power of the tests $(1$-$\beta)$ were recorded as response variables. The simulation study was performed by coding in STATA 13.0.

A Monte Carlo study is conducted to study size and power properties. The powers of the methods are computed in 3000 repetitions under the alternative hypothesis of the existence of break in at least one of constant term and slope coefficient. The break is introduced by taking different cases of error term variance. Also, the break set at different locations in the sample with different sample sizes.

Factors		Number of factor levels	Values of factor levels		
			Level 1	Level 2	Level 3
Difference in constant terms (dc)	$|\beta_{01}-\beta_{02}|$	3	0.00	0.50	1.00
Difference in slope coefficients (ds)	$|\beta_{11}-\beta_{12}|$	3	0.00	0.50	1.00
Break-point (bp)	n_1/n	3	0.10	0.25	0.50
Sample size (ss)	n	2	60	120	
Difference in variance of error terms (de)	$|\sigma_{t1}^2-\sigma_{t2}^2|$	2	0.0	0.3	

Table 3.1 Factors of experiment.

Table 3.2 shows the power of the tests on the variation of the constant term depending on the location of the break-point and sample size without any variability in slope coefficient and error variance. In general, while *ds* and *dt* are invariant, it can be said that as *dc*, *bp* and *ss* increases, the power of the tests performs better. In other words, a bigger break that occurs later in a larger sample can be detected much better by tests. For example, when $dc=1.0$, $bp=0.1$ and $ss=60$, 5 of the 10 methods (*CBT*, *HCT*, *DCT*–1, *HPT* and *CST*) showed high performance, meanwhile $dc=1.0$, $bp=0.5$ and $ss=120$, 8 of the 10 methods (*CBT*, *CPT*, *HCT*, *DCT*-1, *HPT*, *CST*, *CSQT* and *MZ*) showed very strong performance.

METHODS	ss																	
	60									120								
	bp									bp								
	0.1			0.25			0.5			0.1			0.25			0.5		
	dc			dc			dc			dc			dc			dc		
	0.0	0.5	1.0	0.0	0.5	1.0	0.0	0.5	1.0	0.0	0.5	1.0	0.0	0.5	1.0	0.0	0.5	1.0
CBT	4.60	51.20	**98.30**	5.47	84.20	**100.00**	5.53	91.67	**100.00**	4.83	83.57	**99.97**	5.03	99.10	**100.00**	4.97	99.97	**100.00**
CPT	4.87	6.00	8.93	4.47	11.20	40.30	5.87	25.53	**90.83**	5.67	6.97	14.33	5.20	16.97	71.43	4.60	46.27	**99.73**
HCT	7.00	44.73	**96.97**	5.13	83.07	**100.00**	5.03	90.93	**100.00**	6.33	84.87	**99.97**	5.57	99.23	**100.00**	4.63	99.97	**100.00**
DCT–1	5.40	57.63	**99.07**	5.27	90.23	**100.00**	4.73	95.57	**100.00**	4.53	89.00	**99.97**	5.00	99.63	**100.00**	5.87	99.97	**100.00**
DCT–2	4.63	4.63	4.63	4.77	4.77	4.77	5.33	5.33	5.33	4.60	4.60	4.60	5.03	5.03	5.03	4.70	4.70	4.70
HPT	7.80	43.23	**95.90**	11.00	66.90	**99.80**	22.83	80.63	**99.97**	7.00	59.60	**99.87**	9.03	84.57	**100.00**	16.87	93.77	**100.00**
RRT	0.50	2.40	23.47	0.50	2.40	23.47	0.50	2.40	23.47	0.50	2.40	23.47	0.50	2.40	23.47	0.50	2.40	23.47
CST	3.40	84.37	**100.00**	3.37	84.37	**100.00**	3.37	84.37	**100.00**	3.37	84.37	**100.00**	3.37	84.37	**100.00**	3.37	84.37	**100.00**
CSQT	0.57	5.83	85.83	0.57	5.80	85.83	0.57	5.80	85.83	7.60	27.83	**98.60**	7.60	27.83	**98.60**	7.60	27.83	**98.60**
MZ	0.27	8.13	69.30	0.07	33.03	**98.83**	0.37	49.80	**99.93**	0.27	31.67	**98.67**	0.33	75.57	**100.00**	0.23	90.27	**100.00**

Table 3.2 The powers of tests in the $ds=0.0$ and $de=0.0$ cases

The powers for $ds=0.0, 0.5, 1.0$ with different values of *bp* and *ss* while $dc=0.0$ and $de=0.0$ cases are shown in Table 3.3. As in Table 3.2, it can be said that in a larger sample, the bigger break that occurs later is better detected by the methods. With the model that $dc=0.0$, $de=0.0$, $ds=0.1$ with different *bp* and different *ss*, some of the tests like as *CBT*, *HCT*, *DCT*–2, *HPT* and *CSQT* gives the similar high powers. *DCT*–1 test is a method that is sensitive to the variability of the constant term, so it does not work much for the variability in the slope

coefficient. On the other hand, it can be said that the *CST* and the *MZ* tests perform slightly worse under these conditions.

METHODS	ss																	
	60									120								
	bp									bp								
	0.1			0.25			0.5			0.1			0.25			0.5		
	ds			ds			ds			ds			ds			ds		
	0.0	0.5	1.0	0.0	0.5	1.0	0.0	0.5	1.0	0.0	0.5	1.0	0.0	0.5	1.0	0.0	0.5	1.0
CBT	4.60	53.30	**98.90**	5.47	81.33	**100.00**	5.53	87.57	100.00	4.83	83.57	100.00	5.03	99.57	100.00	4.97	100.00	**100.00**
CPT	4.87	5.77	9.10	4.47	10.37	36.90	5.87	23.07	85.90	5.67	7.10	14.30	5.20	18.53	75.60	4.60	49.97	**99.97**
HCT	7.00	67.33	**99.60**	5.13	86.57	100.00	5.03	87.33	100.00	6.33	92.47	**100.00**	5.57	99.77	**100.00**	4.63	100.00	100.00
DCT–1	5.40	5.40	5.40	5.27	5.27	5.27	4.73	4.73	4.73	4.53	4.53	4.53	5.00	5.00	5.00	5.87	5.87	5.87
DCT–2	4.63	59.90	**99.53**	4.77	88.17	**100.00**	5.33	93.40	100.00	4.60	89.47	**100.00**	5.03	99.87	**100.00**	4.70	100.00	**100.00**
HPT	7.80	45.03	**97.10**	11.00	63.87	**99.83**	22.83	81.77	100.00	7.00	59.67	**99.80**	9.03	87.33	**100.00**	16.87	92.20	**100.00**
RRT	0.50	3.03	56.90	0.50	3.03	56.90	0.50	3.03	56.90	0.50	3.03	56.90	0.50	3.03	56.90	0.50	3.03	56.90
CST	3.40	3.07	2.33	3.37	3.07	2.33	3.37	3.07	2.33	3.37	3.07	2.33	3.37	3.07	2.33	3.37	3.07	2.33
CSQT	0.57	5.83	**91.33**	0.57	5.80	**91.33**	0.57	5.80	**91.33**	7.60	31.70	**99.53**	7.60	31.70	**99.53**	7.60	31.70	**99.53**
MZ	0.27	0.33	2.20	0.07	1.63	7.27	0.37	2.23	15.80	0.27	0.87	8.60	0.33	3.43	36.27	0.23	5.67	44.47

Table 3.3 The powers of tests in the *dc*=0.0 and *de*=0.0 cases

METHODS	ss																	
	60									120								
	bp									bp								
	0.1			0.25			0.5			0.1			0.25			0.5		
	dc			dc			dc			dc			dc			dc		
	0.0	0.5	1.0	0.0	0.5	1.0	0.0	0.5	1.0	0.0	0.5	1.0	0.0	0.5	1.0	0.0	0.5	1.0
CBT	49.83	**92.73**	100.00	25.97	**98.67**	100.00	5.07	**99.10**	100.00	46.13	**99.43**	100.00	25.47	**99.97**	100.00	5.83	**100.00**	100.00
CPT	0.23	0.63	2.23	0.00	0.10	9.40	0.00	0.57	62.53	0.00	0.03	0.60	0.00	0.00	12.30	0.00	0.17	94.13
HCT	40.97	**90.80**	**99.93**	24.57	**98.60**	100.00	6.83	**99.50**	100.00	44.90	**99.33**	100.00	25.13	**99.97**	100.00	4.73	**100.00**	100.00
DCT–1	34.03	**90.03**	**100.00**	20.33	**99.03**	100.00	5.17	**99.73**	100.00	33.23	**99.13**	100.00	19.03	**100.00**	100.00	5.20	**100.00**	100.00
DCT–2	34.80	34.80	34.80	17.77	17.77	17.77	3.53	3.53	3.53	32.20	32.20	32.20	19.13	19.13	19.13	6.00	6.00	6.00
HPT	91.50	**99.20**	100.00	99.63	**99.97**	100.00	100.00	**100.00**	100.00	98.87	**100.00**	100.00	100.00	**100.00**	100.00	100.00	**100.00**	100.00
RRT	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
CST	12.17	**97.00**	100.00	12.17	**97.00**	100.00	12.17	**97.00**	100.00	12.17	**97.00**	100.00	12.17	**97.00**	100.00	12.17	**97.00**	100.00
CSQT	99.93	12.17	37.43	99.93	12.17	37.43	99.93	12.17	37.43	99.97	42.37	80.20	99.97	42.37	80.20	99.97	42.37	80.20
MZ	14.33	65.90	**99.30**	31.37	**95.63**	100.00	31.93	**99.50**	100.00	38.00	**96.07**	100.00	80.73	**100.00**	100.00	90.57	**100.00**	100.00

Table 3.4 The powers of tests in the *ds*=0.0 and *de*=0.3 cases

METHODS	ss																	
	60									120								
	bp									bp								
	0.1			0.25			0.5			0.1			0.25			0.5		
	ds			ds			ds			ds			ds			ds		
	0.0	0.5	1.0	0.0	0.5	1.0	0.0	0.5	1.0	0.0	0.5	1.0	0.0	0.5	1.0	0.0	0.5	1.0
CBT	49.83	**94.03**	**100.00**	25.97	**98.13**	**100.00**	5.07	**99.27**	100.00	46.13	**99.03**	100.00	25.47	**100.00**	100.00	5.83	**100.00**	100.00
CPT	0.23	0.50	2.17	0.00	0.13	7.47	0.00	0.57	49.57	0.00	0.03	0.57	0.00	0.00	17.60	0.00	0.33	96.63
HCT	40.97	**93.97**	**100.00**	24.57	**98.17**	**100.00**	6.83	**99.37**	100.00	44.90	**99.27**	100.00	25.13	**100.00**	100.00	4.73	**100.00**	100.00
DCT-1	34.03	34.03	34.03	20.33	20.33	20.33	5.17	5.17	5.17	33.23	33.23	33.23	19.03	19.03	19.03	5.20	5.20	5.20
DCT-2	34.80	**92.00**	**99.97**	17.77	**98.90**	**100.00**	3.53	**99.73**	100.00	32.20	**98.93**	100.00	19.13	**100.00**	100.00	6.00	**100.00**	100.00
HPT	91.50	**99.23**	**100.00**	99.63	**100.00**	**100.00**	100.00	**100.00**	100.00	98.87	**100.00**	100.00	100.00	**100.00**	100.00	100.00	**100.00**	100.00
RRT	0.10	0.10	1.07	0.10	0.10	1.07	0.10	0.10	1.07	0.10	0.10	1.07	0.10	0.10	1.07	0.10	0.10	1.07
CST	12.17	7.17	2.67	12.17	7.17	2.67	12.17	7.17	2.67	12.17	7.17	2.67	12.17	7.17	2.67	12.17	7.17	2.67
CSQT	99.93	15.93	52.50	99.93	15.93	52.50	99.93	15.93	52.50	99.97	49.23	88.37	99.97	49.23	88.37	99.97	49.23	88.37
MZ	14.33	20.83	46.93	31.37	47.40	80.00	31.93	53.60	**90.30**	38.00	55.63	80.50	80.73	**91.93**	**99.17**	90.57	**96.90**	**99.93**

Table 3.5 The powers of tests in the *dc*=0.0 and *de*=0.3 cases

When $de=0.3$ and $ds=0.0$, the powers in cases with $dc=0.0, 0.5, 1.0$ are considered with the different break-point locations and different sample sizes are provided in Table 3.4. Here, since there is a break in both the constant term and the error variance, it is seen that the differences between the powers of the tests are more evident. Also, even with a change of 0.5 units in the constant term, the methods perform very well. This table shows that *CBT*, *HCT*, *DCT*–1, *HPT*, *CST* and *MZ* tests have better power overall in these cases.

Table 3.5 gives the powers of the tests for the models that $dc=0.0$, $de=0.3$ with different ds, bp and ss. Here, it is seen that fewer methods are stronger according to the Table 3.4. According to Table 3.5, *CBT*, *HCT*, *DCT*–2 and *HPT* work very well under all possible cases that $dc=0.0$, $de=0.3$ and $ds=0.0, 0.5, 1.0$. In addition, in some cases where $dc=0.0$ and $de=0.3$ the *MZ* test also performed very well. For example, in the $ds=0.5$ or 1.0, $ss=120$ and $bp=0.25$ or 0.50 cases, the powers of the *MZ* test are higher.

4. CONCLUSION AND DISCUSSION

In this paper, we have considered methods for determining structural break in the linear regression model with cross-section data and compared them in terms of power performances. Different results were obtained from the ten methods applied on regression models in different forms on the data sets produced in the same conditions. Our simulations show that different methods have quite different behavior under the various alternative. These differences may be caused by the assumptions of the methods.

These findings were obtained as a result of the analysis. It is observed that the power of the tests increases with a longer duration of the sample since a shift rather late in the sample period means that each test has much time to determine.

The power of the *MZ* test is higher in $de=0.3$ cases than *in* $de=0.0$ cases. Because the *MZ* test considers with the break in any coefficients in the heteroscedastic case. The *MZ* test is more important than the other tests in that it deals with the change in error term variance and the change in coefficients simultaneously.

Since *DCT*–1 test sensitive to the break in a constant term, it shows high performance in $ds=0.0$ cases. Similarly, *DCT*–2 is a method that sensitive to the break in slope coefficient and showed high performance in $dc=0.0$ cases.

As *CPT* is a method based on using the big sample, it can be inferred that the break-point has a statistically significant effect on *CPT*. It can also be inferred that the place of the break-point change according to before and after the break. For example, from Table 3.2, in the $dc=1.0$, $bp=0.5$ and $ss=120$ case, the power of the *CPT* is 99.73, while in the $dc=1.0$, $bp=0.5$ and $ss=60$ case it is 90.83.

It is seen that from all tables the size of the sample (number of observations) is not effective on *RRT*, *CST*, and *CSQT*. Therefore, the number of observations has no distinctive feature on the power of these tests.

At the end of the study, it was determined that the powers of these methods differ for the structural break based on the cross-sectional data structure. To make better measurements and take healthier results, it may be suggested to apply more than one tests on data structure, rather than applying only one test.

REFERENCES

Ahmed, M., G. Haider, and A. Zaman, (2017). Detecting Structural Change with Heteroscedasticity. *Communications in Statistics-Theory and Methods,* 46, 10446-10455.

Antoshin, S., A. Berg and M. Souto (2008). Testing for Structural Breaks in Small Samples. *IMF Working Paper* 75.

Brown, R.L., J. Durbin and J.M. Evans (1975). Techniques for Testing the Constancy of Regression Relationships over Time. *Journal of the Royal Statistical Society,* 37, 149-192.

Chow, C. Gregory (1960/3). Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica,* 28, 591-605.

Gujarati, D.N. (1970). Use of Dummy Variables in Testing For Equality between Sets of Coefficients in Linear Regressions: A Generalization. *The American Statistician,* 24 (5), 18-22.

Gujarati, D.N. (2004). *Basic Econometrics* (4th Edition). New York City: McGraw-Hill Companies.

Guris, S. and E. Caglayan (2010). *Ekonometri Temel Kavramlar* [Basic Econometric Notions]. Istanbul: Der Yayınevi.

Hansen, E.B. (1992/4). Testing For Parameter Instability in Linear Models. *Journal of Policy Modeling,* 14, 517-533.

Hendry, D.F. (1989). PC-Give: An Interactive Econometric Modeling System. The *University of Oxford. Hendry, D. F. (1989).* Institute of economics and statistics and Nuffield College: University of Oxford.

Maasoumi, E., A. Zaman and M. Ahmed (2010). Testing for Structural Change, Aggregation, and Homogeneity. *Economic Modeling,* 27 (6), 1382-1391.

Toyoda, T. (1974/3). Use of the Chow Test under Heteroscedasticity. *Econometrica,* 42, 601-608.