

Majune, Socrates Kraidó

**Preprint**

## Theory and Practice of Testing for a Single Structural Break in Stata

*Suggested Citation:* Majune, Socrates Kraidó (2018) : Theory and Practice of Testing for a Single Structural Break in Stata, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/172517>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# **Theory and Practice of Testing for a Single Structural Break in Stata**

**Socrates Kraido Majune\***

**January, 2018**

## **Abstract**

The major objective of this paper is to demonstrate, theoretically and empirically, the test of a single structural break/change. Failure to address a structural break can lead to forecasting errors and the general unreliability of a model. Three approaches of testing for structural change are discussed using data from Johnston et al. (1997, p.130) on Stata 14 software. The first approach assesses whether there is a structural break in parameters (slope and intercept) while the second and third assess whether there is a break in slope and intercept respectively. The Residual Sum of Squares (RSS) for the restricted and unrestricted models are established to necessitate the use of an F-test in making inferences. According to the first approach, a structural break exists at 5% level of significance. This result is confirmed by the Chow test. The second and third approaches establish that the structural break is from the intercept and not the slope. These results are also affirmed by the Chow test. Furthermore, all these results, from the first to the third approach, are confirmed by an alternative approach which relies on the knowledge that  $t^2 = F$ . Therefore, the dependent variable is not affected by the policy change on the explanatory variable but it is mainly affected by the basic unobserved qualitative characteristics of the two sub-periods. For further analysis, it is recommended that a unit root test be conducted using the Zivot-Andrews test. This test has been established as the panacea for the interplay between unit root and structural changes.

JEL Classification: C01; C52; C80; C87

Key words: Structural Break, Chow Test, Unit Root, Zivot-Andrews Test

---

\* Department of Economics, University of Dar es Salaam, Tanzania. E-mail: mkraidosocrates@gmail.com

## 1.0 Introduction

A Structural Change/Break is a change or a shift in the common operations of an economy. It might be caused by *inter alia*, external forces such as the world financial crisis of 2008-2009, policy changes such as capping of interest rates to four percent above Central Bank's lending rates in Kenya in 2016, and economic crises. Hence, the main interest of Structural Breaks is that the values of the parameters of the model do not remain constant through the entire time period (Gujarati & Porter, 2009). Hence, it is important to check for Structural Changes especially in time-series data.

This paper focuses on a single structural break caused by a policy change. As a result, three different approaches of testing for this Structural Break are discussed. Thereafter, alternative approaches of testing for the Structural Break are discussed. These discussions involve theoretical and practical illustrations based on Johnston & DiNardo (1998, pp. 126-133) and Stata 14 software is used.

This paper is organized as follows. After this brief introduction in section 1, the nature of data used is discussed in section 2. Thereafter, approaches of structural breaks are extensively discussed in theory and empirics in section 3. Section 4 is on alternative approaches while section 5 is the conclusion.

## 2.0 Data

This section discusses how data is uploaded in Stata 14 and how other variables are generated. A link to the data, and the Do-File are provided in the appendix.

Original data from Johnston & DiNardo (1998, p. 130) is of the form:

$$y_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 4 \\ 6 \end{bmatrix} \quad x_1 = \begin{bmatrix} 2 \\ 4 \\ 6 \\ 10 \\ 13 \end{bmatrix} \quad y_2 = \begin{bmatrix} 1 \\ 3 \\ 3 \\ 5 \\ 6 \\ 6 \\ 7 \\ 9 \\ 9 \\ 11 \end{bmatrix} \quad x_2 = \begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \\ 10 \\ 12 \\ 14 \\ 16 \\ 18 \\ 20 \end{bmatrix}$$

To enter this data from the Keyboard, the input command is used followed by  $x$  and  $y^2$  as follows<sup>3</sup>: *input x y*. Afterwards, a time variable is created using the command *gen time=\_n*. Time helps to partition data as it will be seen shortly. Thereafter, two dummies are generated based on time where dummy1 has period 1 to 5 as having structural break ( $d1=1$ ) while the period after 5 does not have structural break ( $d1=0$ ). Equally, dummy 2 has period 1 to 5 having no structural break ( $d2=0$ ) while the period after 5 has structural break ( $d2=1$ ). The subsequent commands are: *gen d1=1* and *replace d1=0 if time >=6* for dummy 1 and; *gen d2=1* and *replace d2=0 if time <=5* for dummy 2. The slope depicting the break is generated by the command *gen Z1=x* followed by *replace Z1=0 if time >=6*. Similarly, the slope for the post-break period is generated by the command *gen Z2=x* followed by *replace Z2=0 if time <=5*. The final data is presented in Table 1<sup>4</sup>.

**Table 1: Data**

x	y	time	d1	d2	Z1	Z2
2	1	1	After the Break	Before the Break	2	0
4	2	2	After the Break	Before the Break	4	0
6	2	3	After the Break	Before the Break	6	0
10	4	4	After the Break	Before the Break	10	0
13	6	5	After the Break	Before the Break	13	0
2	1	6	Before the Break	After the Break	0	2
4	3	7	Before the Break	After the Break	0	4
6	3	8	Before the Break	After the Break	0	6
8	5	9	Before the Break	After the Break	0	8
10	6	10	Before the Break	After the Break	0	10
12	6	11	Before the Break	After the Break	0	12
14	7	12	Before the Break	After the Break	0	14
16	9	13	Before the Break	After the Break	0	16
18	9	14	Before the Break	After the Break	0	18
20	11	15	Before the Break	After the Break	0	20

<sup>2</sup> This is regardless of whether the variable is  $x_1$  or  $x_2$  and  $y_1$  or  $y_2$ .

<sup>3</sup> The usually steps of preparing Stata for data were done e.g. *set more off* and labeling data. See Do-File for details.

<sup>4</sup> Also see Figure 1 in the appendix.

*Note: After the Break=1 while Before the Break=0. A command, logout, save(list) word replace fix: list, uploads results directly into word.*

### 3.0 Analysis of Structural Breaks

This section discusses three approaches used to determine the presence of a structural break and its source. Each approach involves establishment of Residual Sum of Squares (RSS) from a restricted and an unrestricted model which are encompassed in an F-statistic<sup>5</sup>. The restricted model assumes that the sub-period regressions are not different and that the intercept and the slope remain the same for the entire period (Gujarati & Porter, 2009). Conversely, the unrestricted model assumes that the intercept and the slope coefficients are different for the sub-periods (Gujarati & Porter, 2009). Thereafter, inference is made by comparing the F-statistic with the F-critical value. Each approach is discussed below alongside their respective Chow test which is an alternative approach.

#### 3.1 The Common Approach (Tests of slope and intercept)

The break is assumed to arise from both the intercept and the slope (all parameters) in this case. Given that the break is only for two periods, period 1 (pre) and period 2 (post) which occurs after the break, we let  $y_i, x_i$  ( $i = 1, 2$ ) to indicate the appropriate partitioning of the data. Similarly, we let  $n = n_1 + n_2$ , where  $n_1$  and  $n_2$  are the number of observations for period 1 and period 2 respectively.

Therefore, the resulting restricted model is of the form:

$$\begin{bmatrix} y_{pre} \\ y_{post} \end{bmatrix} = \begin{bmatrix} i_{pre} & x_{pre}^* \\ i_{post} & x_{post}^* \end{bmatrix} \begin{bmatrix} \alpha \\ \beta^* \end{bmatrix} + u \quad u \sim N(0, \sigma^2 I) \dots\dots\dots 1$$

The subsequent unrestricted model is of the form:

$$\begin{bmatrix} y_{pre} \\ y_{post} \end{bmatrix} = \begin{bmatrix} i_{pre} & 0 & x_{pre}^* & 0 \\ 0 & i_{post} & 0 & x_{post}^* \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1^* \\ \beta_2^* \end{bmatrix} + u \quad u \sim N(0, \sigma^2 I) \dots\dots\dots 2$$

The restricted model in (1) assumes that the intercept as well as the slope coefficient remain the same over the entire period in that  $\alpha = \alpha_1 = \alpha_2$  and  $\beta^* = \beta_1^* = \beta_2^*$ . Its degrees of freedom are  $n_1 + n_2 - k$  or  $n - k$  where  $k$  is the number of parameters being estimated.

---

<sup>5</sup>  $F = \frac{(RSS_R - RSS_{UR})/df_*}{RSS_{UR}/df} \sim F(df_*, df)$

The unrestricted model in (2) assumes that the intercept and the slope coefficients are different. Hence each period is regressed differently to yield the following degrees of freedom;  $n_1 - k$  for period 1 and  $n_2 - k$  for period 2. Adding the two degrees of freedom generates the degrees of freedom for the unrestricted model as follows:  $n_1 - k + n_2 - k = n_1 + n_2 - 2k = n - 2k$ . The next step involves getting the Residual Sum of Squares for the restricted model i.e.  $RSS_R$  and the unrestricted model  $RSS_{UR} = RSS_1 + RSS_2$ .

Thereafter, the F-statistic is obtained as follows:

$$F = \frac{(RSS_R - RSS_{UR})/k}{(RSS_{UR})/(n-2k)} \sim F_{[k, (n-2k)]}$$

The degrees of freedom for the numerator is established as follows:  $(RSS_R - RSS_{UR}) = n - k - (n - 2k) = k$ .  $u$  in (1) and (2) represents the error term.

The null and alternative hypotheses are as follows:

**Null hypothesis:**  $H_0: \alpha_1 = \alpha_2, \beta_1^* = \beta_2^*$  (There is no break in the intercept and slope)

**Alternative hypothesis:**  $H_1: \alpha_1 \neq \alpha_2, \beta_1^* \neq \beta_2^*$  (There is a break in the intercept and slope)

Therefore, the resulting F calculated value is tested against the F critical value from the table. The null hypothesis is rejected if the F calculated value is greater than the F critical value and vice versa.

Results of the restricted model are presented in Table 2. A command; *reg y x* is used to obtain results.

**Table 2: Regression results for the Restricted Model**

. reg y x						
Source	SS	df	MS	Number of obs	=	15
Model	127.443885	1	127.443885	F(1, 13)	=	252.71
Residual	6.55611511	13	.504316547	Prob > F	=	0.0000
				R-squared	=	0.9511
				Adj R-squared	=	0.9473
Total	134	14	9.57142857	Root MSE	=	.71015
y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x	.5244604	.0329917	15.90	0.000	.4531862	.5957347
_cons	-.0697842	.3678736	-0.19	0.852	-.8645268	.7249584

This has an RSS of 6.55611511 with 13 (15-2) degrees of freedom.

Results of the unrestricted model are presented in Table 3 prior to the command; *reg y d1 d2 Z1 Z2*.

**Table 3: Regression results for the Unrestricted Model**

<pre>. reg y d1 d2 Z1 Z2 note: d1 omitted because of collinearity</pre>						
Source	SS	df	MS	Number of obs	=	15
Model	130.839773	3	43.6132576	F(3, 11)	=	151.81
Residual	3.16022727	11	.287293388	Prob > F	=	0.0000
				R-squared	=	0.9764
				Adj R-squared	=	0.9700
Total	134	14	9.57142857	Root MSE	=	.536

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
d1	0	(omitted)				
d2	.4625	.6062146	0.76	0.462	-.8717693	1.796769
Z1	.4375	.0599263	7.30	0.000	.305603	.569397
Z2	.5090909	.0295057	17.25	0.000	.4441493	.5740325
_cons	-.0625	.4831417	-0.13	0.899	-1.125888	1.000888

Results in Table 3 indicate that the RSS is 3.16022727 with 11 (15-4) degrees of freedom.

Hence, the F- calculated value becomes;

$$F = \frac{(RSS_R - RSS_{UR})/k}{(RSS_{UR})/(n-2k)} = F = \frac{(6.55611511 - 3.16022727)/2}{(3.16022727)/(11)} = 5.91$$

Comparing this calculated value to the critical value at 5% level of significance ( $F_{[2,11]} = 3.9823$ ), the null hypothesis is rejected and the conclusion is that there is a structural break/change. However, this result is rejected at 1% level of significance because the critical value,  $F_{[2,11]} = 7.20571$ , is greater than the calculated value.

The **Chow test**<sup>6</sup> is used as an alternative test of structural change. We first regress a model with all the variables, *reg y x d1 d2 Z1 Z2*, then test for equality of the coefficients of dummies and slope separately which yields an F-statistic. Results are as shown in Table 4:

**Table 4: Results for Chow test (First Approach)**

<pre>. test _b[ d1 ]=_b[ d2 ], notest</pre>			
( 1)	o.d1 - d2 = 0		
<pre>. test _b[ Z1 ]=_b[ Z2 ], accum</pre>			
( 1)	o.d1 - d2 = 0		
( 2)	Z1 - Z2 = 0		
	F( 2, 11) =	5.91	
	Prob > F =	0.0181	

<sup>6</sup> Check (Gujarati & Porter, 2009, pp. 254-257) and (Stata, 2017) for more details.

Results in Table 4 yield an F-Statistic similar to the earlier calculated one of 5.91. From the prior inference, there is a structural break at 5% level of significance but there is no structural break when testing at 1% level of significance.

Next, we illustrate a second approach of conducting a Chow Test. In this approach, there is need to have correct knowledge of when the structural break occurred, which is period 6<sup>7</sup>. Hence, the command used is as follows: *estat sbknown, break(6)*. This should be preceded by a declaration of data as time series using the command: *tsset time*. Results are presented in Table 5.

**Table 5: Results for Chow test (Second Approach)**

. estat sbknown, break(6)			
Wald test for a structural break: Known break date			
	Number of obs	=	15
Sample:	1 - 15		
Break date:	6		
Ho: No structural break			
	chi2(2)	=	11.8203
	Prob > chi2	=	0.0027
Exogenous variables:	x		
Coefficients included in test:	x _cons		

From Table 5, it can be concluded that there is a structural break as earlier determined. This is done using the Chi-square test in which the calculated value, 8.5407, is greater than the critical, 5.99146, at 5% level of significance with two degrees of freedom. Hence, the null hypothesis is rejected and the conclusion is that there is a structural break.

### 3.2 Tests of Slope Coefficients

In this case, the restriction is on the slope other than the intercept. Interest is in establishing whether the break arises from the slope. Similar to the discussion in section 3.1, we let  $y_i, x_i$  ( $i = 1, 2$ ) to indicate the appropriate partitioning of the data in period 1 and period 2. Furthermore, we let  $n = n_1 + n_2$ , where  $n_1$  and  $n_2$  are the number of observations for the period 1 and period 2 respectively.

The resulting restricted model is of the form:

$$\begin{bmatrix} y_{pre} \\ y_{post} \end{bmatrix} = \begin{bmatrix} i_{pre} & 0 & x_{pre}^* \\ 0 & i_{post} & x_{post}^* \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta^* \end{bmatrix} + u \quad u \sim N(0, \sigma^2 I) \dots \dots \dots .3$$

<sup>7</sup> See trend in of X in Figure 1 in the Appendix.



Its degrees of freedom are  $n - k - 1$ .

The corresponding unrestricted model is of the form:

$$\begin{bmatrix} y_{pre} \\ y_{post} \end{bmatrix} = \begin{bmatrix} i_{pre} & 0 & x_{pre}^* & 0 \\ 0 & i_{post} & 0 & x_{post}^* \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1^* \\ \beta_2^* \end{bmatrix} + u \quad u \sim N(0, \sigma^2 I) \dots \dots \dots 4$$

Alternatively, the unrestricted model can also be written as:

$$\begin{bmatrix} y_{pre} \\ y_{post} \end{bmatrix} = \begin{bmatrix} i_{pre} & 0 & x_{pre}^* & 0 \\ 0 & i_{post} & 0 & x_{post}^* \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 - \alpha_1 \\ \beta_1^* \\ \beta_2^* - \beta_1^* \end{bmatrix} + u \quad u \sim N(0, \sigma^2 I) \dots \dots \dots 5$$

Synonymous to the model in section 3.1, its degrees of freedom are  $n - 2k$ .

The resulting F-statistic is calculated as follows:

$$F = \frac{(RSS_R - RSS_{UR}) / (k - 1)}{(RSS_{UR}) / (n - 2k)} \sim F_{[(k-1), (n-2k)]}$$

Where the degrees of freedom for the numerator is established as follows:  $(RSS_R - RSS_{UR}) = n - k - 1 - (n - 2k) = k - 1$ .  $u$  in (3), (4) and (5) represents the error term.

This is compared to the F value from the table at 5% level of significance. Notably, the null and alternative hypotheses are:

$H_0: \beta_1^* = \beta_2^*$  (There is no break in the slope)

$H_1: \beta_1^* \neq \beta_2^*$  (There is a break in the slope)

Results of the restricted model, shown in Table 6, are established using the command *reg y dl x*.

**Table 6: Regression results for the Restricted Model**

. reg y dl x						
Source	SS	df	MS	Number of obs	=	15
Model	130.509756	2	65.254878	F(2, 12)	=	224.36
Residual	3.4902439	12	.290853659	Prob > F	=	0.0000
Total	134	14	9.57142857	R-squared	=	0.9740
				Adj R-squared	=	0.9696
				Root MSE	=	.53931
y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
dl	-1.019512	.3140167	-3.25	0.007	-1.703696	-.3353286
x	.495122	.0266345	18.59	0.000	.4370903	.5531536
_cons	.5536585	.3390025	1.63	0.128	-.1849644	1.292281

RSS from Table 6 is 3.4902439 with 12 degrees of freedom.

The corresponding results of the unrestricted model are presented in Table 7. They are obtained using the command: *reg y d1 d2 Z1 Z2*.

**Table 7: Regression results for the Unrestricted Model**

<pre>. reg y d1 d2 Z1 Z2 note: d1 omitted because of collinearity</pre>						
Source	SS	df	MS	Number of obs	=	15
Model	130.839773	3	43.6132576	F(3, 11)	=	151.81
Residual	3.16022727	11	.287293388	Prob > F	=	0.0000
				R-squared	=	0.9764
				Adj R-squared	=	0.9700
Total	134	14	9.57142857	Root MSE	=	.536
y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
d1	0	(omitted)				
d2	.4625	.6062146	0.76	0.462	-.8717693	1.796769
Z1	.4375	.0599263	7.30	0.000	.305603	.569397
Z2	.5090909	.0295057	17.25	0.000	.4441493	.5740325
_cons	-.0625	.4831417	-0.13	0.899	-1.125888	1.000888

RSS from Table 7 is 3.16022727 with 11 degrees of freedom.

Therefore, F- calculated value becomes;

$$F = \frac{(RSS_R - RSS_{UR}) / (k - 1)}{(RSS_{UR}) / (n - 2k)} = F = \frac{(3.4902439 - 3.16022727) / 1}{(3.16022727) / (11)} = 1.15$$

This is compared to F-critical from the table,  $F_{[1,11]} = 4.84434$ , at 5% level of significance.

Given that the calculated value is less than the critical value,  $1.15 < 4.84434$ , the null hypothesis is not rejected and the conclusion is that there is no structural break/change arising from the slope.

This can alternatively be tested using the Chow test similar to section 3.1. The appropriate command is *test \_b[ Z1 ]=\_b[ Z2 ]* and it yields a priori F-statistic of 1.15 as shown in Table 8.

**Table 8: Chow test Results for Test of Slope Coefficients**

<pre>. test _b[ Z1 ]=_b[ Z2 ]</pre>			
( 1)	Z1 - Z2 = 0		
	F( 1, 11) =	1.15	
	Prob > F =	0.3068	

The F-Calculated value from Table 8 is 1.15. As earlier noted, this indicates that there is no structural break arising from the slope.

### 3.3 Tests of Intercepts

In this case, interest is to determine whether the structural break arises from the intercepts in that intercepts from regression equations intersect the y axis at the same point. Furthermore, since the

slope is assumed to be common, the main aim is to determine between the restricted and unrestricted model, which is higher or lower than the other in the direction of the dependent variable.

From prior discussions, we let  $y_i, x_i$  ( $i = 1, 2$ ) to indicate the appropriate partitioning of the data in period 1 and period 2. Furthermore, we let  $n = n_1 + n_2$ , where  $n_1$  and  $n_2$  are the number of observations for the period 1 and period 2 respectively.

The ensuing restricted model is of the form:

$$\begin{bmatrix} y_{pre} \\ y_{post} \end{bmatrix} = \begin{bmatrix} i_{pre} & x_{pre}^* \\ i_{post} & x_{post}^* \end{bmatrix} \begin{bmatrix} \alpha \\ \beta^* \end{bmatrix} + u \quad u \sim N(0, \sigma^2 I) \dots \dots \dots .6$$

With  $n - k$  degrees of freedom.

The ensuing unrestricted model is of the form:

$$\begin{bmatrix} y_{pre} \\ y_{post} \end{bmatrix} = \begin{bmatrix} i_{pre} & 0 & x_{pre}^* \\ i_{post} & i_{post} & x_{post}^* \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 - \alpha_1 \\ \beta^* \end{bmatrix} + u \quad u \sim N(0, \sigma^2 I) \dots \dots \dots .7$$

With  $n - k - 1$  degrees of freedom.

The resulting F-statistic is calculated as follows:

$$F = \frac{(RSS_R - RSS_{UR}) / (1)}{(RSS_{UR}) / (n - k - 1)} \sim F_{[1, (n - k - 1)]}$$

Where the degrees of freedom for the numerator is established as follows:  $(RSS_R - RSS_{UR}) = n - k - (n - k - 1) = 1$ .  $u$  in (6) and (7) represents the error term.

This is compared to the F value from the table at 5% level of significance. The null and alternative hypotheses are:

$H_0: \alpha_1 = \alpha_2$  (There is no break in the intercept)

$H_1: \alpha_1 \neq \alpha_2$  (There is a break in the intercept)

The command: *reg y x* is used to establish results of the restricted model in Table 9.

**Table 9: Regression results for the Restricted Model**

. reg y x						
Source	SS	df	MS	Number of obs	=	15
Model	127.443885	1	127.443885	F(1, 13)	=	252.71
Residual	6.55611511	13	.504316547	Prob > F	=	0.0000
				R-squared	=	0.9511
				Adj R-squared	=	0.9473
Total	134	14	9.57142857	Root MSE	=	.71015
y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x	.5244604	.0329917	15.90	0.000	.4531862	.5957347
_cons	-.0697842	.3678736	-0.19	0.852	-.8645268	.7249584

RSS from Table 9 is 6.55611511 with 13 (15-2) degree of freedom.

Results of the corresponding unrestricted model are presented in Table 10 using the command:

*reg y d2 x.*

**Table 10: Regression results for the Unrestricted Model**

. reg y d2 x						
Source	SS	df	MS	Number of obs	=	15
Model	130.509756	2	65.254878	F(2, 12)	=	224.36
Residual	3.4902439	12	.290853659	Prob > F	=	0.0000
				R-squared	=	0.9740
				Adj R-squared	=	0.9696
Total	134	14	9.57142857	Root MSE	=	.53931
y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
d2	1.019512	.3140167	3.25	0.007	.3353286	1.703696
x	.495122	.0266345	18.59	0.000	.4370903	.5531536
_cons	-.4658537	.3048463	-1.53	0.152	-1.130057	.1983495

RSS in this case is 3.4902439 with 12 degrees of freedom.

Therefore, the F- calculated value becomes;

$$F = \frac{(RSS_R - RSS_{UR})/(1)}{(RSS_{UR})/(n-k-1)} = F = \frac{(6.55611511 - 3.4902439)/1}{(3.4902439)/(12)} = 10.54$$

This is compared to F-critical from the table,  $F_{[1,12]} = 4.74723$ , at 5% level of significance and

$F_{[1,12]} = 9.33021$  at 1% level of significance.

Since the calculated value is greater than the critical value in both cases,  $10.54 > 4.74723$  and  $10.54 > 9.33021$ , the null hypothesis is rejected and the conclusion is that there is a structural break arising from the intercept.

However, we were unable to establish an appropriate command for the Chow test in this case.

#### 4.0 Alternative approaches

An alternative approach can be used to get the results in section 3. This alternative approach involves using the relationship between the t-statistic and the F-statistic in which  $t^2 = F$  (Greene, 2012, p. 160).

The first illustration is based on approaches 3.1 and 3.2 which are the Common Approach (Tests of slope and intercept) and Tests of Slope Coefficients. The respective unrestricted model is shown in (8) and it is similar to (5) which is also an unrestricted model.

$$\begin{bmatrix} y_{pre} \\ y_{post} \end{bmatrix} = \begin{bmatrix} i_{pre} & 0 & x_{pre}^* & 0 \\ 0 & i_{post} & 0 & x_{post}^* \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 - \alpha_1 \\ \beta_1^* \\ \beta_2^* - \beta_1^* \end{bmatrix} + u \quad u \sim N(0, \sigma^2 I) \dots\dots\dots 8$$

The command for this equation is: `reg y x d2 Z2`.

**Table 11: Regression results for the Unrestricted Model (Alternative Approach)**

. reg y x d2 Z2						
Source	SS	df	MS	Number of obs = 15		
Model	130.839773	3	43.6132576	F(3, 11)	=	151.81
Residual	3.16022727	11	.287293388	Prob > F	=	0.0000
Total	134	14	9.57142857	R-squared	=	0.9764
				Adj R-squared	=	0.9700
				Root MSE	=	.536
y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x	.4375	.0599263	7.30	0.000	.305603	.569397
d2	.4625	.6062146	0.76	0.462	-.8717693	1.796769
Z2	.0715909	.0667964	1.07	0.307	-.0754269	.2186087
_cons	-.0625	.4831417	-0.13	0.899	-1.125888	1.000888

The RSS in Table 11 is 3.16022727. Comparing it with the RSS of the ideal restricted model in (1) and (6), 6.55611511, yields an F statistics of 5.91. That is:

$$F = \frac{(RSS_R - RSS_{UR})/k}{(RSS_{UR})/(n-2k)} = F = \frac{(6.55611511 - 3.16022727)/2}{(3.16022727)/(11)} = 5.91$$

This is compared to the F-critical value,  $F_{[2,11]} = 3.9823$ , at 5% level of significance. Since the calculated value is greater than the critical value,  $5.91 > 3.9823$ , the null hypothesis is rejected and the conclusion is that there is a structural break. However, there no structural break when tested at 1% level of significance because  $5.91 < 7.20571$ .

The Tests of Slope Coefficients is also established using the t-statistic. The intention is to test whether Z2 is significant. Taking its t-value of 1.07 and squaring it gives yields a value of 1.15.

This is similar to the F-Statistic in section 3.2. As earlier established, the conclusion is that there is no structural break arising from the slope.

To establish the results of Tests of Intercepts similar to section 3.3, (7)<sup>8</sup> is used. Results of the corresponding unrestricted model are as follows:

**Table 12: Regression results for the Unrestricted Model (Alternative Approach)**

. reg y d2 x						
Source	SS	df	MS	Number of obs	=	15
Model	130.509756	2	65.254878	F(2, 12)	=	224.36
Residual	3.4902439	12	.290853659	Prob > F	=	0.0000
				R-squared	=	0.9740
				Adj R-squared	=	0.9696
Total	134	14	9.57142857	Root MSE	=	.53931
y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
d2	1.019512	.3140167	3.25	0.007	.3353286	1.703696
x	.495122	.0266345	18.59	0.000	.4370903	.5531536
_cons	-.4658537	.3048463	-1.53	0.152	-1.130057	.1983495

The point of interest here is to examine the significance of d2 whose t-statistic is 3.25. Squaring this value gives 10.54, which as earlier established in section 3.3, leads to the conclusion of a structural break arising from the intercept.

## 5.0 Conclusion

The major objective of this paper was to demonstrate the test of a single structural break/change. This assessment was done in threefold: First, we determined whether a structural break existed (arising from slope and intercept); second we determined whether the break was from the slope and; third we determined whether the break was from the intercept. Using data from Johnston and DiNardo (1998, p. 130), both theoretical and empirical arguments were discussed with the latter argument using Stata 14. All the three cases involved calculating an F-statistic arising from the ratio of Residual Sum of Squares (RSS) of restricted and unrestricted models. Restricted models assume that sub-period regressions are not different and that the intercept and the slope remain the same for the entire period while unrestricted models recognize the difference in intercept and slope

<sup>8</sup>(7) is of the form: 
$$\begin{bmatrix} y_{pre} \\ y_{post} \end{bmatrix} = \begin{bmatrix} i_{pre} & 0 & x_{pre}^* \\ i_{post} & i_{post} & x_{post}^* \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 - \alpha_1 \\ \beta^* \end{bmatrix} + u \text{ where } u \sim N(0, \sigma^2 I)$$

over time. The final inference was made from comparing the F-statistic with the F-critical value. These results were then confirmed in two ways; first using an inherent Stata command of Chow test and then an alternative approach with the knowledge that  $t^2 = F$ .

The first approach confirmed the presence of a structural break in both the intercept and the slope when tested at 5% level of significance. However, there was no structural break when tested at 1% level of significance. This finding was affirmed by the Chow test and later on by the alternative approach in section 4. Based on the inference of a structural break in the data (using 5% level of significance), the second approach tested whether the break was from the slope. It was established that the break did not arise from the slope. This was confirmed by the Chow test and later on in section 4. The third option tested whether this break was from the intercept. Inference from the F-test indicated that indeed the break was from the intercept. This was also confirmed by the alternative approach of section 4. Therefore, the dependent variable was mainly affected by the basic unobserved qualitative characteristics of the two sub-periods and not the policy which led to changes in explanatory variable.

It is recommended that further analysis should be done to test for unit root using the Zivot-Andrews test (Andrews & Zivot, 1992). It has been found to be the most suitable test of addressing the interplay between structural change and unit root. Further analysis should also explore cases of multiple breaks. Lastly, the strength of this study is on the availability of a detailed do-file and a list of selected studies on structural breaks in the appendix.

## References

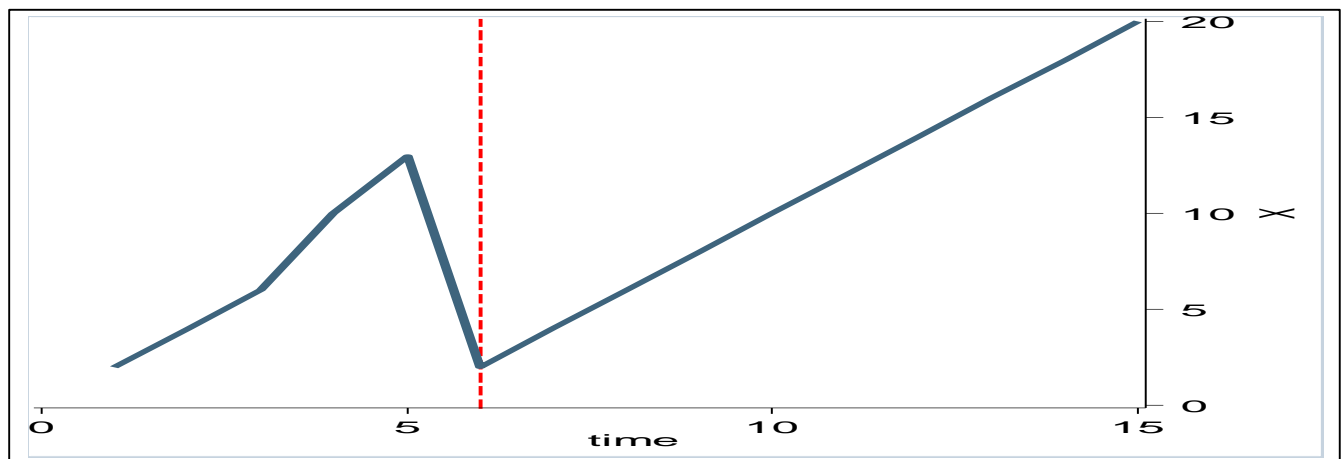
- Allaro, H. B., Kassa, B., & Hundie, B. (2011). A time series analysis of structural break time in the macroeconomic variables in Ethiopia. *African Journal of Agricultural Research*, 6(2), 392-400.
- Andrews, D., & Zivot, E. (1992). Further evidence on the Great Crash, the oil price shock, and the unit-root hypothesis. *Journal of Business and Economic Statistics*, 10, 251- 270.
- Greene, W. (2012). *Econometric Analysis* . Essex: Pearson Education Limited.
- Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics* . New York: McGraw-Hill/Irwin.
- Johnston, J., & DiNardo, J. (1998). *Econometric Methods* (4 ed.). Singapore:: McGraw-Hill.

- Mussagy, I. H., & Allaro, H. (2016). Structural Change in Mozambique: Economic Performance Before and After the Civil War. *Journal of Economics and Sustainable Development*, 7(11), 119-125.
- Nair, M. S., Narayana, N., & Rathedi, M. (2013). Structural Break and Economic Diversification in Botswana. *Asian-African Journal of Economics and Econometrics*, 13(1), 85-97.
- Ndirangu, L., García, C., and Gitau, C. (2014). Evidence of Structural Breaks in Macroeconomic Variables in Kenya. Paper prepared for the 2014 CSAE Conference to be held on 23rd to 25th March 2014, UK. Retrieved from [https://editorialexpress.com/cgi-bin/conference/download.cgi?db\\_name=CSAE2014&paper\\_id=511](https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=CSAE2014&paper_id=511).
- Ndoricimpa, A. (2013). Structural breaks and fiscal deficit sustainability in EAC countries: Empirical evidence. *International Journal of Economics, Finance and Management Sciences*, 1(6), 391-399.
- Ruranga, C., Ocaya, B., & Kaberuka, W. (2014). Analysis of Rwandan Economic Performance Before and After the 1994 Genocide. *African Journal of Economic Review*, 2(2), 1-18.
- Stata. (2017, July 1). *Chow test* . Retrieved from Stata:  
<http://www.stata.com/support/faqs/statistics/chow-tests>

## APPENDIX

### APPENDIX 1: Trend of X

**Figure 1: Trend of X**



Note: Vertical line indicates estimated break location



## APPENDIX 2: Link to Data

<https://drive.google.com/file/d/0B1sKLvZeHqXhR2dZOUFrdDJ4UXM/view?usp=sharing>

## APPENDIX 3: Do-File

```
set more off
//To enter data manually//
input x y
2 1
4 2
6 2
10 4
13 6
2 1
3 4
6 3
8 5
10 6
12 6
14 7
16 9
18 9
20 11
end
//End helps to finalize the data entry//
edit
save dummyProfKidane
replace x = 4 in 7
replace y = 3 in 7
//Generate time variable//
gen time=_n
help label
```

```

label data "Dummy data from Johnstone et al(1997)"
describe
label variable x "X variable"
label variable y "Y variable"
label variable time "Generated time variable"
//To see how the data looks//
list
//Generate dummy 1//
gen d1=1
tab d1
replace d1=0 if time >=6
list d1
gen d2=1
label variable d1 "Dummy 1"
replace d2=0 if time<=5
list d2
label variable d2 "Dummy 2"
label define d1 0 "Before the Break" 1 "After the Break"
label values d1 d1
tab d1
label define d2 0 "Before the Break" 1 "After the Break"
label values d2 d2
//Genare z1 and z2 as of slope before the break. Z1 values of Xs before the break and Z2 after
the break//
gen Z1=x
replace Z1=0 if time>=6
list
gen Z2=x
replace Z2=0 if time<=5
list Z2
//To see how the whole dataset looks, use list//

```

```

list
//This result can directly be uploaded into word using the command below//
logout, save(list) word replace fix: list
//Use page 131 to generate tables//
//To obtain results of the restricted model run the following regression//
reg y x
//To obtain results for the unrestricted model, use this command//
reg y d1 d2 Z1 Z2
//Compare the calculated values with critical F- value of 5.91. Use the residual and the
formulas. This is to make an inference on structural break//
//Use the values of the residual in the formula//
//Alternatively, a Chow test can be conducted after declaring data as time series//
tsset time
//First Approach of the Chow test//
//Run a regression with all the variables//
reg y x d1 d2 Z1 Z2
//Test for the equality of the coefficients of dummies and Zs//
test _b[ d1 ]=_b[ d2 ], notest
test _b[ Z1 ]=_b[ Z2 ], accum
//Interpret the value of the F statistic in comparison to the F critical value from the table//
//Second approach//
//Just run the Chow test command specifying the time of break and interpret the results//
estat sbknown, break(6)
//Then the Chi-Square value is used for interpretation.//
//To test for slope coefficient.//
//The RSS from the restricted model is found using the following command.//
reg y d1 d2 x
//The RSS for the unrestricted model is found using.//
reg y d1 d2 Z1 Z2
//The Chow test in this case is conducted using the formula//
test _b[ Z1 ]=_b[ Z2 ]

```

```
//To test for intercepts.//
//The RSS for restricted model is obtained by//
reg y x
//RSS for the unrestricted model is obtained by//
reg y d2 x
//Alternative approaches//
//First run the
//Alternatively, use the t-test on page 132. to test for structural break. The F-that we got before
should be the same as the square of the t-value//
//Use the t-value. Square the value of Z2. It should give the same F as before (t^2). This tests
for the slope.//
reg y d2 x
//To plot the graph in Figure 1//
twoway (tsline x), xline(6)
log close
```

#### APPENDIX 4: Discussion of selected studies in Africa

**Table 13: List of Selected Studies in Africa on Structural Breaks**

Author (s)	Country	Type of Data	Findings
(Mussagy & Allaro, 2016)	Mozambique	Annual data (1980 – 2014)	Presence of structural break on Gross Domestic Product, Foreign Direct Investment, Gross Capital Formation, Savings and the Exports.
Ndirangu et al. (2014)	Kenya	Monthly (October 1997 to march 2013) and annual data (1973 to 2011)	Presence of structural breaks from unstable political, economic, trade liberalization, poor and erratic climatic conditions and external shocks (International commodity price shocks and global financial crisis.
(Ruranga et al., 2014)	Rwanda	Annual data (1973 – 2011)	Presence of structural breaks in the pre and post-genocide period arising from slope coefficients and not intercept.

(Ndoricimpa, 2013)	EAC	Annual data (1985-2012)	Presence of a structural break in government spending (% GDP) and government revenue (% GDP) among all EAC countries (Burundi, Kenya, Uganda, Tanzania and Rwanda).
(Nair et al., 2013)	Botswana	Annual data (1974 - 2009)	Presence of two breaks i.e. 1987-88 and 1996-97. Agriculture was the only key sector that posted a positive break.
(Allaro et al., 2011)	Ethiopia	Annual data (1974 – 2009)	Presence of a structural break on GDP, exports and imports in 2003. This was 11 years after the regime shift.