Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 13 (3): 503-507, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.503.507

Likelihood Ratio Type Test and its Critical Values for Structural Change of Unknown Changepoint

[1]Mohammad Nurul Azam, [2]Mahbuba Yeasmin and [3]Gazi Mahabubul Alam
[1]Department of Quantitative Analysis, College of Business Administration,
King Saud University, Riyadh 11451, Kingdom of Saudi Arabia
[2]EKRA Institute, Melbourne, Victoria 3166, Australia
[3]Academic Performance Enhancement Unit, University of Malaya, 50603 Kuala Lumpur, Malaysia

Abstract: This study considers the problem of testing for a structural change of unknown timing in a regression coefficient in the linear regression model. This is a non-standard testing problem and practical important situation facing applied modelers. Simulation methods were used to generate a range of exact critical values of the Likelihood Ratio (LR) type test for different sample sizes, numbers of regressors and types of regressors. We found that the critical values depend on sample size, the number of regressors and to a less extend on the type of explanatory variables. We recommend using the LR type test statistic for testing structural change of unknown timing with our critical value.

Key words: Testing, structural change, changepoint, likelihood ratio statistics, critical value, monte carlo simulation

INTRODUCTION

The presence of a structural change in data that is not detected is a hazard for applied researchers, with serious consequences for model performance and forecasting. The structural Changepoint problems have been studied by many authors (Chernoff and Zacks, 1964; Sen and Srivastava, 1975; Hawkins, 1977; Yao and Davis, 1984; Worsley, 1986; Singh and Pandey, 2011; Jothilakshmi *et al*., 2011; Kamurzzaman and Takeya, 2008). It is helpful to have a test for structural change when the changepoint is unknown. Andrews (1990) compared the Likelihood Ratio (LR) test with tests such as the CUSUM and CUSUM of squares tests and the fluctuation test of Sen (1980) and Ploberger *et al*. (1989) in terms of power. Andrews (1993) determined the asymptotic distributions of the LR test statistics under the null hypothesis of parameter stability. In modern times the likelihood ratio test studied for the simple linear regression models, such as Kim and Siegmund (1989) and Kim (1994). The asymptotic results on likelihood approach for linear models applied (Csorgo and Horvath, 1997). Guan (2004) used empirical likelihood method which was applied to detect the changepoint. There have also been extensive studies on LR test problems those can found (Koroto, 2009; Aziz *et al*., 2011; Midi *et al*., 2011).

In this study, we propose a new approach which is based on likelihood ratio type method to test the change in regression coefficient for linear model in the presence of an unknown changepoint. There are so many ways one can develop a test statistic to test for the presence of structural change when there is a possible unknown changepoint in the data. Since LR test does not have a known distribution for finite sample sizes, we estimate exact critical values for the test by simulation for different sample sizes, numbers of regressors and types of regressors.

THE MODEL

We consider the linear regression model for t = 1,···,n, with a possible change of unknown timing in one coefficient:

$$y_t = \begin{cases} x_t'\beta_0 + w_t\gamma + u_t & \text{for} \quad t < n_1; \\ x_t'\beta_0 + w_t(\gamma+\delta) + u_t & \text{for} \quad t \geq n_1, \end{cases} \quad (1)$$

where, y_t is the dependent variable at time t, x_t is a k×1 vector of regressors at time t, w_t is a scalar variable that is of interest, β_0 is a k×1 vector of regression coefficients and γ and δ are unknown scalar parameters. u_t~N(0, σ^2). Model (1) can be written jointly as:

Corresponding Author: Gazi Mahabubul Alam, Academic Performance Enhancement Unit, University of Malaya, 50603 Kuala Lumpur, Malaysia

$$y_t = x'_t\, \beta_0 + w_t\gamma + z_t\, \delta + u_t, \quad \text{for } t = 1,\cdots,n \tag{2}$$

where, z_t is a dummy variable defined as:

$$z_t = \begin{cases} 0, & t \le n_1, \\ w_t, & t > n_1. \end{cases} \tag{3}$$

Denoting:

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$X_1 = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} & w_1 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} & w_n \end{bmatrix}$$

$$X_2 = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} & w_1 & z_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} & w_n & z_n \end{bmatrix}$$

$$u = [u_1 \cdots u_n]'$$

$$\beta_0 = [\alpha_0 \;\; \beta_1 \cdots \beta_{k-1}]'$$

$$\theta_1 = [\beta_0 \;\; \gamma]'$$

$$\theta_2 = [\beta_0 \;\; \gamma \;\; \delta]'$$

Model (2) can be rewritten in matrix form as:

$$y = \begin{cases} X_1\theta_1 + u, & \text{under } H_0, \\ X_2\theta_2 + u, & \text{under } H_1, \end{cases} \tag{4}$$

Where:

$$u \sim N(0, \sigma^2 I_n)$$

TEST STATISTICS

The null hypothesis of interest is:

$$H_0: \delta = 0$$

and the alternative hypothesis is:

$$H_1: \delta \ne 0$$

The loglikelihood function of the sample under the alternative hypothesis that there is a changepoint in the data after period n_1 is:

$$l_1 = -\frac{n}{2}\log(2\pi\sigma_2^2) - \frac{1}{2\sigma_2^2}(y - X_2\,\theta_2)(y - X_2\,\theta_2) \tag{5}$$

The loglikelihood function under the null hypothesis of no changepoint in the data is:

$$l_0 = -\frac{n}{2}\log(2\pi\sigma_1^2) - \frac{1}{2\sigma_1^2}(y - X_1\,\theta_1)(y - X_1\,\theta_1) \tag{6}$$

Differentiating (5) with respect to the parameters β_0, γ, δ and σ^2_2 and equating the resultant equations to zero, we obtain the conditional Maximum Likelihood (ML) estimates of β_0, γ, δ and σ^2_2 under the alternative hypothesis that there is a changepoint in the data after period n_1 as:

$$\hat{\theta}_2 = (X_2'X_2)^{-1}X_2'y, \qquad \hat{\sigma}_2^2 = (y - X_2\hat{\theta}_2)'(y - X_2\hat{\theta}_2)/n \tag{7}$$

Differentiating (6) with respect to the parameters β_0, γ and σ^2_1 and equating the resultant equations to zero, we obtain the ML estimates of β_0, γ and σ^2_1 under the null hypothesis that there is no changepoint in the data as:

$$\hat{\theta}_1 = (X_1'X_1)^{-1}X_1'y, \qquad \hat{\sigma}_1^2 = (y - X_1\hat{\theta}_1)'(y - X_1\hat{\theta}_1)/n \tag{8}$$

Substituting the estimates of (7) into (5), we obtain the concentrated log-likelihood function under the alternative hypothesis of the sample given a changepoint in the data after period n_1 as:

$$\hat{l}_1 = -\frac{n}{2}\log(2\pi\hat{\sigma}_2^2) - \frac{n}{2} \tag{9}$$

Substituting the estimated values from (8) into (6), we obtain the maximized log-likelihood function under the null hypothesis of no changepoint:

$$\hat{l}_0 = -\frac{n}{2}\log(2\pi\hat{\sigma}_1^2) - \frac{n}{2} \tag{10}$$

When the changepoint n_1 is unknown, a naturally, intuitive approach would be to estimate n_1 and then apply the LR test at that estimate of n_1. Given n_1, (9) is maximized by substituting in the estimated value of σ^2_2 from (7). Maximizing (9) with respect to n_1 is equivalent to finding the n_1 for which $\hat{\sigma}^2_2$ is minimum. The LR test statistic can be obtained by substituting minimum values of $\hat{\sigma}^2_2$ in (9) which we will denote by $\hat{\hat{l}}_1$ and then taking twice the difference between it and the log-likelihood of (10) that is, $LR = 2(\hat{\hat{l}}_1 - \hat{l}_0)$.

SIMULATIONS

The data y_t were generated from the following equation:

$$y_t = x_t' \beta_0 + w_t \gamma + u_t, \text{ for } t = 1, \cdots, n$$

where, $u_t \sim N(0,1)$, w_t is a scalar variable, β_0 is a k×1 vector of regression coefficients and γ is a constant coefficient. The k×1 independent variables are generated following (Watson and Engle, 1985). Monte Carlo experimental design; that is explanatory variables (excluding the constant term) were generated from the first order autoregressive process $x_{it} = \phi x_{it-1} + e_{it}$ with $e_{it} \sim IN(0,1)$ for $t = 1, \cdots, n$, where ϕ takes values 0, 0.7, 1.0 and 1.02 which covers white noise, autoregressive, random walk and explosive processes, respectively. The number of regressors k was allowed to range from 1 to 15 in turn with ϕ being the same for each regressor and w_t was generated from the uniform distribution with range from 0 to 1. Four different sample sizes of 25, 50, 75 and 100 were used. We generated 50,000 LR test statistics for each set of ϕ's, s, k and n, then ordered the calculated LR from the lowest to the highest values and obtained the 90th, 95th, 97th and 99th percentiles which are the required critical values for the 10, 5, 2.5 and 1% level of significance, respectively. Throughout, we use the GAUSS for Windows Version 3.2.35 software to estimate the parameters of the model by the method of ML estimation. In the model, error terms were simulated using pseudo random numbers from the GAUSS function RNDNS that generates standard normal variate for regression errors. The seed for the random number generator for each experiment was 1786.

RESULTS OF THE SIMULATION

Table 1 report the critical value calculation results of the Monte Carlo simulations. We discuss the overall trends in the critical values in four stages. The first stage

Table 1: Empirical critical values of the LR test for different numbers of regressors and ϕ

α	1%				2.5%			
k	ϕ = 0	ϕ = 0.7	ϕ = 1.0	ϕ = 1.02	ϕ = 0	ϕ = 0.7	ϕ=1.0	ϕ = 1.02
n = 25								
1	4.779	4.766	4.759	4.837	3.814	3.840	3.829	4.022
5	6.082	6.058	6.441	6.732	4.778	4.811	5.217	5.416
10	8.710	8.652	9.454	9.129	7.006	7.004	7.651	7.590
15	16.196	16.169	17.106	17.453	12.477	12.567	14.090	13.903
5%	**10%**							
1	3.136	3.139	3.179	3.356	2.446	2.455	2.463	2.636
5	3.938	3.962	4.298	4.435	3.031	3.016	3.331	3.514
10	5.478	5.485	6.285	6.299	4.026	4.020	4.715	4.877
15	10.209	10.182	11.249	11.386	7.698	7.735	9.033	8.995
n = 50								
1	5.044	5.046	5.153	5.259	4.237	4.243	4.328	4.504
5	5.350	5.689	5.775	5.881	4.603	4.677	4.932	4.993
10	6.438	6.489	6.785	6.674	5.440	5.512	5.701	5.673
15	7.232	7.263	7.643	7.787	6.166	6.140	6.602	6.568
5%	**10%**							
1	3.570	3.511	3.665	3.791	2.820	2.829	2.980	3.129
5	3.824	3.995	4.212	4.267	3.112	3.152	3.367	3.516
10	4.697	4.682	4.875	4.813	3.724	3.741	4.048	4.034
15	5.280	5.296	5.621	5.587	4.300	4.355	4.678	4.689
n = 75								
1	5.153	5.140	5.228	5.293	4.168	4.161	4.313	4.464
5	5.353	5.344	5.660	5.700	4.458	4.439	4.686	4.760
10	5.764	5.803	6.158	6.266	4.757	4.734	5.134	5.250
15	6.457	6.468	6.569	6.911	5.212	5.180	5.597	5.741
5%	**10%**							
1	3.486	3.484	3.609	3.780	2.783	2.787	2.962	3.062
5	3.683	3.674	3.933	4.077	2.968	2.968	3.186	3.310
10	4.051	4.043	4.389	4.465	3.210	3.189	3.582	3.665
15	4.304	4.284	4.697	4.872	3.450	3.468	3.863	3.957
n = 100								
1	4.906	4.917	4.829	5.302	4.059	4.060	4.105	4.451
5	5.284	5.281	5.658	5.659	4.444	4.424	4.656	4.824
10	5.678	5.744	5.981	6.071	4.696	4.715	5.068	5.171
15	6.017	6.072	6.610	6.633	4.989	5.029	5.543	5.615
5%	**10%**							
1	3.497	3.511	3.508	3.800	2.818	2.826	2.863	3.129
5	3.757	3.777	3.972	4.046	3.016	3.016	3.310	3.340
10	3.994	4.004	4.304	4.403	3.237	3.249	3.516	3.639
15	4.219	4.232	4.644	4.662	3.448	3.450	3.820	3.838

involves the patterns or trends with respect to sample size variation, the second involves patterns as the number of regressors in the model changes, the third considers changes in the type of autoregressive regressors and the fourth discusses some general patterns with regard to the significance level.

A noticeable feature is that the simulated critical values of the LR test increase as the sample size increases from 25 to 75 for small k = 1 and 2 and it decreases as the sample size increases from 75 to 100 at the 1% level of significance for different values of ϕ considered. At the 2.5, 5 and 10% levels of significance, critical values of the LR test increase as the sample size increases from 25 to 50 and they decrease as the sample size increases from 50 to 75. They decrease as n increases from 75 to 100 for the different values of ϕ considered. The critical values of the LR test increase as the sample size increases for k = 3 or more at different levels of significance for different values of ϕ.

The largest calculated critical value of the test occurs at the 1% level of significance when ϕ = 1.02 and n = 25 and takes the value 17.453 whereas, when n = 100 it takes the value 6.633 The largest critical values of the test at the 2.5, 5 and 10% levels of significance occur when ϕ = 1.02 and n = 25 and are 13.903, 11.386 and 8.995, respectively, whereas, when n = 100, the critical values are respectively, 5.615, 4.662 and 3.838. The minimum critical values of the test statistic at the 1, 2.5, 5 and 10% levels of significance occur when ϕ = 0 and n = 25 and are, respectively, 4.779, 3.814, 3.136 and 2.446, whereas, when n = 100 these critical values are respectively, 4.906, 4.059, 3.497 and 2.818.

The critical values of the test almost always increase with an increase in the number of regressors k. The largest increases occur for small n and for small ϕ. The smallest increases occur for n = 100 and for large levels of significance. For a large number of regressors in the model when ϕ = 0 and k = 15 at the 1, 2.5, 5 and 10% levels of significance, the maximum value of the critical values are 16.196, 12.477, 10.209 and 7.698, respectively.

The simulated critical values of the LR type test statistic appear to be practically unchanged as the type of autoregressive regressors change with everything else held constant. The critical values are typically the same for ϕ = 0 and for ϕ = 0.70 at different levels of significance and are also roughly the same for ϕ = 1.0 and ϕ = 1.02. The latter are almost always slightly bigger than the former. Obviously the critical values decrease as the level of significance increases. This decrease is largest for large k. This variation in critical values with n, k, ϕ and α suggests the need for formulae for critical values of the test statistic which will be developed in the next section.

CONCLUSION

While there are so many ways to develop a test statistic to test for the presence of structural change when there is a possible unknown changepoint in the data, we recommend the use of the LR test. Since this test does not have a known distribution for finite sample sizes, we estimated exact critical values for the test by simulation using 50,000 replications for different sample sizes, numbers of regressors and types of regressors. We found that the critical values clearly depend on sample size, the number of regressors and to a less extend on the type of explanatory variables. Further research direction is how to make inference about the changepoint n_1 when parameters of the models have to be estimated.

REFERENCES

Andrews, D.W.K., 1990. Tests for parameter instability and structural change with unknown change point. Cowles Foundation Discussion Paper 943, Yale University.

Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica, 61: 821-856.

Aziz, D., S.S. Siraj, S.K. Daud, J.M. Panandam and M.F. Othman, 2011. Genetic diversity of wild and cultured populations of *Panaeus monodon* using microsatellite markers. J. Fish. Aquat. Sci., 6: 614-623.

Chernoff, H. and S. Zacks, 1964. Estimating the current mean of a normal distribution which is subjected to changes in time. Ann. Math. Stat., 35: 999-1018.

Csorgo, M. and L. Horvath, 1997. Limit Theorems in Change-Point Analysis. John Wiley and Sons Inc., New York, USA., ISBN-13: 9780471955221, Pages: 414.

Guan, Z., 2004. A semiparametric changepoint model. Biometrika, 91: 849-862.

Hawkins, D.M., 1977. Testing a sequence of observations for a shift in location. J. Am. Statist. Assoc., 72: 180-186.

Jothilakshmi, M., D. Thirunavukkarasu and N.K. Sudeepkumar, 2011. Structural changes in livestock service delivery system: A case study of India. Asian J. Agric. Res., 5: 98-108.

Kamurzzaman, M. and H. Takeya, 2008. Importance, structural change and factors affecting production of vegetables in Bangladesh. Int. J. Agric. Res., 3: 397-406.

Kim, H.J. and D. Siegmund, 1989. The likelihood ratio test for a change-point in simple linear regression. Biometrika, 76: 409-423.

Kim, H.J., 1994. Likelihood ratio and cumulative sum tests for a change-point in linear regression. J. Multivariate Anal., 51: 54-70.

Koroto, T., 2009. An approximate likelihood ratio method for testing equality of two dependent proportions. Asian J. Math. Stat., 2: 15-19.

Midi, H., S.K. Sarkar and S. Rana, 2011. Adequacy of multinomial logit model with nominal responses over binary logit model. Trends Applied Sci. Res., 6: 900-909.

Ploberger, W., W. Kramer and K. Kontrus, 1989. A new test for structural stability in the linear regression model. J. Econ., 40: 307-318.

Sen, A.K. and M.S. Srivastava, 1975. On tests for detecting changes in mean. Ann. Statist., 3: 98-108.

Sen, P.K., 1980. Asymptotic theory of some tests for a possible change in the regression slope occurring at an unknown time point. Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebeite, 52: 203-218.

Singh, N.J. and R.K. Pandey, 2011. Rehydration characteristics and structural changes of sweet potato cubes after dehydration. Am. J. Food Technol., 6: 709-716.

Watson, M.W. and R.F. Engle, 1985. Testing for regression coefficient stability with a stationary AR1 alternative. Rev. Econ. Statist., 67: 341-346.

Worsley, K.J., 1986. Confidence regions and tests for a change-point in a sequence of exponential family random variables. Biometrika, 73: 91-104.

Yao, Y.C. and R.A. Davis, 1984. The asymptotic behavior of the likelihood ratio statistic for testing a shift in mean in a sequence of independent normal variate. Sankhya Ser. A, 48: 339 -353.