



2014.02.01.STAT.02

JOINPOINT REGRESSION ANALYSIS AND AN APPLICATION ON ISTANBUL STOCK-EXCHANGE*

Huriye TELLİ*

Sinan SARAÇLI†

Afyon Kocatepe University, Faculty of Arts and Sciences, Department of Statistics, Afyonkarahisar

Received: 24 March 2014

Accepted: 12 May 2014

Özet

Joinpoint Regresyon Analizi, trendde meydana gelen istatistiksel olarak anlamlı değişmelerin en uygun noktalarını belirlemede kullanılan istatistiksel metotlardan biridir. Bu çalışmanın amacı Joinpoint regresyon analizini borsa verilerine uygulamak ve bu metodun performansını gerçek ve tahmin edilen değerleri karşılaştırarak belirlemektir. Bu amaçla Nisan-Mayıs 2013 ayları için Ulusal BIST 30 endeks değerlerine ilişkin veri seti derlenerek Joinpoint analizi ile incelenmiştir. Joinpoint paket programı aracılığı ile doğrusal ve doğrusal olmayan tekniklere göre çözümlene yapılarak bu tekniklerden hangisinin daha uygun olduğu Hata Kareler Ortalaması (HKO) değerlerine göre belirlenmiştir. Gelecek dönemler için yapılan tahminler ile bu dönemlerde gerçekleşen değerlere göre tahmin edilen değerlerin gerçek değerlerden biraz yüksek olduğu gözlemlenmiştir ama bu sonuç Joinpoint regresyon analizinin gelecek dönem tahminlerini gerçekleştirmek için bir zaman serisi verisine uygulanabileceğini göstermektedir.

Anahtar Kelimeler: BIST 30, Joinpoint Regresyon Analizi, Trend

Jel Kodu: C1, C4, C5.

Abstract

Joinpoint Regression Analysis is one of the statistical methods used to identify the best-fitting points if there is a statistically significant change in the trend. The aim of this study is to apply joinpoint regression analysis in the stock market and compare the performance of this method according to actual data set and estimated values. For this purpose, we collected the data set from the National Istanbul Stock Exchange (ISE) 30 index for April-May 2013 and examined that data set via Joinpoint Regression Analysis. We applied linear and nonlinear techniques with the help of Joinpoint software and determined the best technique according to their Mean Square Errors (MSE). With the projection for the future months and the actual results, we see that the estimated values are a little higher than the actual values. However, this shows that we may apply Joinpoint regression to a time series data set in order to forecast future values.

Keywords: Istanbul Stock-Exchange 30, Joinpoint Regression Analysis, Trend

Jel Code: C1, C4, C5.

* This study is a part of Huriye Telli's MS thesis supervised by Sinan Saraçlı at Afyon Kocatepe University Institute of Science, 2013.

* huriyetelli@hotmail.com

† ssaracli@aku.edu.tr

1. Introduction

Regression Analysis is a method that models the relationship between the dependent and the independent variable(s) with mathematical expressions. This relationship between variables is actually a causal relationship expressed by a mathematical function.

Regression analysis is applied to estimate the average value of the dependent variable by the help of the data set of the all variables and to test whether or not the independent variables in the model are statistically significant on the dependent variable (Çankal, 2010).

Like many other statistical methods, regression analysis needs to meet some assumptions. Least Squares Technique is one of the most used techniques to estimate the parameters in regression analysis (Gazeloğlu, 2012). This technique is also suitable to use in the case of the assumption of constant variance (Albayrak, 2008).

In Regression Analysis numbers of independent variables may be one, and often two or more. Hence, if this relationship between the variables is linear and the number of independent variables is one, this model is called "Simple Linear Regression Model". Simple linear regression model is given in Equation 1. (Vural, 2007)

$$Y_i = \beta_0 + \beta_1 X_i + e_i \quad (1)$$

The Joinpoint Analysis was used to identify the best-fitting points where a statistically significant change in the trend occurred (Cayuela *et al.*, 2004). Joinpoint Regression, also known as piecewise linear regression, is increasingly used to identify the timing and extent of changes in time series (Goovaerts and Xiao, 2011). Segmented line regression, also known in the literature as multiphase regression with the continuity constraint, broken line regression, piecewise linear regression, has been successfully applied to describe these trend data, and various statistical techniques have been developed for inference problems (Kim *et al.*, 2004).

2. Material and Method

The Joinpoint Regression Software is a Windows-based statistical software package that computes and analyzes non-linear, piecewise trends of time series. The software fits models to the data that allow for testing of whether an apparent change in trend is statistically significant. These models are usually referred to as joinpoint models. It is important to understand that goal of the Joinpoint Regression Program is not to provide models that best fit the data, but models that best summarize the behavior or the data trend (Marrett, 2010).

There are five steps involved in generating any Joinpoint trend analysis;

Step 1: Create an input data file for Joinpoint,

Step 2: Set parameters in Joinpoint,

Step 3: Execute Joinpoint,

Step 4: View Joinpoint results, print models, export graphs and data,

Step 5: Summarize Joinpoint results for interpretation (Marrett, 2010).

The identification of changes in the recent trend is an important issue in the analysis (Kim *et al.*, 2000). Simple Linear Joinpoint Regression Model is given in Equation 2.

$$E[y_i | x_i] = \beta_0 + \beta_1 \cdot x_i + \gamma_1(x_i - \tau_1)^+ \\ = + \dots + \gamma_n(x_i - \tau_k)^+ \quad (2)$$

x_i , $i = 1, 2, \dots, n$; $x_1 < \dots < x_n$ time variable, y_i , $i = 1, 2, \dots, n$ are the response variable. β_0 shows constant coefficient in Equation 2 and β_1 shows slope coefficient.

In modeling trends over time, it is important to be able to detect when statistically significant changes in the trend occur. Joinpoint analysis is widely applied to detect these changes points (joinpoints) and determine the trends between joinpoints (Jiang *et al.*, 2010).

Nonlinear Joinpoint Regression Model is given in Equation 3.

$$E[y_i | x_i] = e^{\beta_0 + \beta_1 \cdot x_i + \gamma_1(x_i - \tau_1)^+ + \dots + \gamma_n(x_i - \tau_k)^+} \quad (3)$$

$\beta_0, \beta_1, \gamma_1, \dots, \gamma_n$ are regression coefficients in Equation 2 and 3, τ_k ; $k < n$, is the k -th unknown Joinpoints (Kim et al., 2000). Equation applies to eliminate negative values during the prediction of the observed values in Equation 4. If $(x_i - \tau_k)$ is negative, this value is zero and if $(x_i - \tau_k)$ is positive, this value is accepted as $(x_i - \tau_k)$.

$$(x_i - \tau_i) = \begin{cases} > 0 & (x_i - \tau_k)^+ = (x_i - \tau_k) \\ \leq 0 & \text{otherwise} \end{cases} \quad (4)$$

This model assumes a linear trend between joinpoints and continuity at the joinpoints. The joinpoint regression model has the same underlying assumptions as simple regression. If the year(s) when changes in the trend occur (joinpoints) are known, then linear regression techniques can be used to estimate the regression parameters. However, in most instances, the exact years for the joinpoints are unobservable.

The challenge in cancer trend analysis is to determine the locations of the joinpoints if they exist; and to determine the optimal number of joinpoints for the most appropriate model (Jiang et al., 2010).

There are three major decisions in any joinpoint analysis;

1. The form of the mean function (Data distribution: Normal or Poisson; Equation: linear or nonlinear),
2. The location of the joinpoints given the number of joinpoints,
3. The optimal joinpoint model (Jiang *et al.*, 2010).

The first step is determined by the form of the data. The next step in fitting the model is to determine the range of the number of joinpoints to be tested; usually between 0 and 9. Then for each given number of joinpoints the location of the joinpoints is determined. To determine the location of the joinpoints, either the Grid Search method or Hudson's method can be applied. The grid search method creates a "grid" of all possible locations for joinpoints specified by the settings, and tests the sum square of errors (SSE) at each one to find the best possible fit. Hudson's method

does a continuous testing between observed values to find the best model. In Hudson's method, each fitted sub-model has a local least square, and the overall least squares of the model is found when a complete curve to be fitted consists of two or more sub models. The third step is to find the optimal model, i.e. the optimal number of joinpoints, and the optimal locations of related joinpoints. (Jiang et al., 2010).

Permutation tests and Bayesian Information Criterion (BIC) test may be applied to determine the most appropriate model in the joinpoint regression (Duyar, 1995). BIC is based on goodness of fit test. Calculation of BIC is given in Equation 5.

$$BIC(k) = \ln \{SSE(k) / Obs\} + \{[2 \cdot (k+1)] / Obs\} \cdot \ln(Obs) \quad (5)$$

Where SSE is the sum of squared errors of the k-JoinPoint regression model, $2 \cdot (k+1)$ is the number of parameters of the k-JoinPoint model and Obs is the number of observations. The k-JoinPoint model with the minimum value of BIC(k) is selected as the final model (Pickle et al., 2007).

The sequential method to determine the optimal model using the permutation test is described below;

k_1 , the maximum number of joinpoints, k_0 the minimum number of joinpoints;

H_0 : k_0 is the best model.

H_1 : k_1 is the best model.

$0 \leq k_0 \leq k_1$, and calculate $\alpha' = \frac{\alpha}{k_1 - k_0}$ for a

given significant level α

2) If $p > \alpha'$, don't reject H_0 and set $k_1 = k_1 - 1$,

If $p \leq \alpha'$ then H_0 is rejected, set $k_0 = k_0 + 1$ and calculate new $\alpha' = \frac{\alpha}{k_1 - k_0}$

3) Go back to do step 1 and 2 until $k_1 = k_0 + 1$ (Jiang et al., 2010).

The formula used to calculate the percent change is given in Equation 6.

$$P.C. = 100 * e^{\beta} - 1 \quad (6)$$

In Equation 6, β is the slope coefficient of each section.

3. Application

In the application part of the study, the daily closing values of the stocks at the 2th quarter of the months of April and May in 2013 for National ISE 30 index was examined to estimate the values for June 2013. The estimated values are also compared with the real ones.

The survey was conducted with a total number of 43 days (21 days for April and 22 days for May in 2013) excluding weekends and public holidays. The data set which obtained from the official site of Istanbul Stock Exchange was compiled in a text file and then transferred to Joinpoint regression software. Analyses are concluded both for linear and nonlinear Joinpoint regression models.

Figure 1. The graphical representation for closing data of National ISE 30 Index in April- May

Whether the changes in the trends of ISE-30 index in April-May 2013 are statistically significant or not, the breaking times and graphs of these trends, model estimates and daily percent changes (DPC) are given in Table 1.

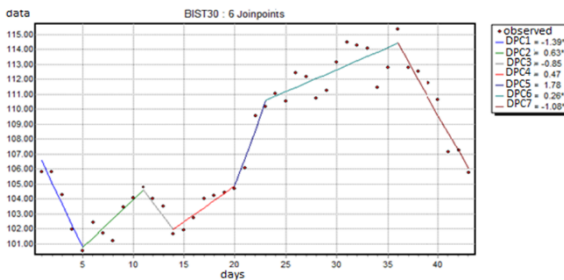


Table 1. Model and the parameters for ISE 30 Index.

Parameter	Value
Model	$\ln(y) = xb / y = xb$
Dependent variable	Daily closing data
Independent variable	Days (Days belonging to

Parameter	Value
	April-May in 2013)
Heteroscedastic errors option	Constant variance
By variable	Groups (ISE 30)
Number of Joinpoints	min=0 ve max=9
Minimum number of observations between two joinpoints	10
Method	Grid Search
Model Selection Method	Bayesian Information Criterion
Significance level	0,05

4. Findings

The scatterplot and the trend lines for National ISE 30 Index for April-May 2013 are given in Figure 1. The results of the analysis show that after some fluctuations in April there is an increase with an amount of 1.78% at the end of the April. After an increase in May with an amount of 0.26% it can be seen that after May 22, there is a decrease with an amount of -1.08% in the total index.

Nonlinear joinpoint Regression Model, according to parameter estimates for April and May is given in Equation 7. The detailed results of the analysis are also given in Table 2.

$$E[y_i | x_i] = e^{(4,68 - 0,01x_i + 0,02(x_i - 5) - 0,01(x_i - 11) + 0,01(x_i - 14) + 0,01(x_i - 20) - 0,02(x_i - 23) - 0,01(x_i - 36))} \quad (7)$$

Table 2. Results for Nonlinear Joinpoint Regression Model

Cohort	Parameter	Standard Parameter		Z value	Probability	MSE	
		Parameter estimates	Standard Error				
ISE30	Constant 1	4.683081	0.011339	412.992596	0.000000	0.00008	
ISE30	Slope 1	-0.014032	0.004141	-3.388868	0.002526		
ISE30	Slope 2- Slope 1	0.020296	0.005071	4.002180	0.000560		
ISE30	Slope 3- Slope 2	-0.014833	0.013417	-1.105525	0.280363		
ISE30	Slope 4- Slope 3	0.013246	0.013417	0.987249	0.333793		
ISE30	Slope 5- Slope 4	0.012979	0.013417	0.967395	0.343410		
ISE30	Slope 6- Slope 5	-0.015035	0.013116	-1.146305	0.263456		
ISE30	Slope 7- Slope 6	-0.013491	0.001913	-7.050813	0.000000		
Test Statistics							
Cohort	Model	Number of Joinpoint	Number of Parameter	Degrees of Freedom	SSE		BIC
ISE 30	#1	0 Joinpoint	2	41	0.0317903	-7.0348560	
ISE 30	#2	1 Joinpoint	4	39	0.0165874	-7.5104305	
ISE 30	#3	2 Joinpoint(s)	6	37	0.0077217	-8.1001004	
ISE 30	#4	3 Joinpoint(s)	8	35	0.0056628	-8.2352841	
ISE 30	#5	4 Joinpoint(s)	10	33	0.0036137	-8.5095150	
ISE 30	#6	5 Joinpoint(s)	12	31	0.0029564	-8.5353500	
ISE 30	#7	6 Joinpoint(s)	14	29	0.0024110	-8.5643233 *	
ISE 30	#8	7 Joinpoint(s)	16	27	0.0022394	-8.4632364	
ISE 30	#9	8 Joinpoint(s)	18	25	0.0020562	-8.3736408	
ISE 30	#10	9 Joinpoint(s)	20	23	0.0018726	-8.2922279	

The selected final model has 6 joinpoint.

The results of Bayesian Information Criterion values indicate that the final selected model has 6 joinpoint.

Linear Joinpoint Regression Model, according to parameter estimates for April and May is given in Equation 8. The detailed results of the analysis are also given in Table 3.

$$\begin{aligned}
 E[y_i | x_i] = & 108,02 - 1,45x_i + 2,10(x_i - 5) \\
 & -1,53(x_i - 11) + 1,36(x_i - 14) \\
 & +1,43(x_i - 20) - 1,61(x_i - 23) - 1,49(x_i - 36)
 \end{aligned}
 \tag{8}$$

Table 3. Results for Linear Joinpoint Regression Model

Cohort	Parameter	Standard Parameter		Z value	Probability	MSE	
		Parameter estimates	Standard error				
ISE 30	Constant 1	108.020953	1.243206	86.888986	0.000000	0.99188	
ISE 30	Slope 1	-1.451426	0.453955	-3.197292	0.004004		
ISE 30	Slope 2- Slope 1	2.096154	0.555979	3.770204	0.000994		
ISE 30	Slope 3- Slope 2	-1.527801	1.470982	-1.038627	0.309775		
ISE 30	Slope 4- Slope 3	1.363997	1.470982	0.927270	0.363415		
ISE 30	Slope 5- Slope 4	1.425030	1.470982	0.968761	0.342742		
ISE 30	Slope 6- Slope 5	-1.612738	1.438039	-1.121484	0.273654		
ISE 30	Slope 7- Slope 6	-1.493279	0.209773	-7.118562	0.000000		
Test Statistics for the number of Joinpoint							
Cohort	Model	Number of Joinpoint	Number of Parameter	Degrees of Freedom	SSE		BIC
ISE 30	#1	0 Joinpoint	2	41	371.3961790	2.3310088	
ISE 30	#2	1 Joinpoint	4	39	190.7380213	1.8395798	
ISE 30	#3	2 Joinpoint(s)	6	37	87.6198435	1.2366260	
ISE 30	#4	3 Joinpoint(s)	8	35	63.895907	1.0958084	
ISE 30	#5	4 Joinpoint(s)	10	33	41.5241008	0.8397716	
ISE 30	#6	5 Joinpoint(s)	12	31	35.4030696	0.8552357	
ISE 30	#7	6 Joinpoint(s)	14	29	28.764525	0.8225205 *	
ISE 30	#8	7 Joinpoint(s)	16	27	26.8953358	0.9302691	
ISE 30	#9	8 Joinpoint(s)	18	25	24.5636540	1.0145236	
ISE 30	#10	9 Joinpoint(s)	20	23	22.1702821	1.0869480	

The selected final model has 6 joinpoint.

Because the MSE value of the nonlinear joinpoint regression analysis is lower than the MSE value of linear model, we can say that the nonlinear joinpoint model is more appropriate to conclude the estimations for June.

Estimations for June 2013 by the help of the data set for April - May are given in Table 4. By

Table 4. Estimates and the actual values for June, 2013 by the help of the data set of April and May in 2013

Cohort	April - May 2013					Actual values (June in 2013)			
	(Estimated values June 3-4-5 2013)								
	y = xb		ln(y) = xb			3	4	5	
ISE 30	3	4	5	3	4	5	3	4	5
	105,02	103,83	102,64	113,30	112,17	111,05	94,83	99,20	97,60

examining Table 4, it can be seen that the actual values are smaller than the estimated values.

5. Conclusion

The results of the analysis indicate that Joinpoint regression analysis can be used to make some estimations for the stock market. The estimated values in this study are similar, but a little higher than the actual values of the related data set.

According to the data set for April-May 2013 there aren't many fluctuations in April. However, there is an increase in May. There is a serious decrease after May 20.

During this period, after the declaration of U.S.A Federal Reserve chairman, that they stop injecting money into the economy, there had been a negative impact on the Istanbul stock market.

Also the Istanbul stock market had been badly affected because of the Gezi Park Protests.

It can be said that there had been a statistically significant decrease in the closing values of traded companies in the stock exchange related with these events.

The estimated values in April are a little higher than the actual data set for closing values of the National ISE 30 index than in May. However, the estimated values in May and June are a little lower than the actual values because of the economic and political events which occurred during these months.

References

1. Albayrak, A. S. (2008). Değişen Varyans Durumunda En Küçük Kareler Tekniğinin Alternatifi Ağırlıklı Regresyon Analizi ve Bir Uygulama. Afyon Kocatepe Üniversitesi İİBF Dergisi, 2: 111-134.
2. Cayuela, A., Dominguez, S.R., Campos, J.L., Candelera, R.O., Matutes, C.R. (2004). Joinpoint Regression Analysis of Lung Cancer Mortality, Andalusia 1975–2000. *Annals of Oncology*, 15: 793-796.
3. Çankal, E. (2010). İstatistik, Lisan Yayınları, Ankara.
4. Duyar, İ. (1995). Model Seçiminde Bayesian Yaklaşım. Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, 2(28): 95-106.
5. Gazeloğlu, C. (2012). Doğrusal Tip II Regresyon Tekniklerinin Monte-Carlo Benzetim Çalışması ile Karşılaştırılması: Sağlık, Bulanık ve Sağlık Bulanık Teknikler. Yüksek Lisans Tezi, Afyon Kocatepe Üniversitesi, Fen Bilimleri Enstitüsü, Afyonkarahisar.
6. Goovaerts, P. and Xiao, H. (2011). Geographical, Temporal and Racial Disparities in Late-Stage Prostate Cancer Incidence Across Florida: A Multiscale Joinpoint Regression Analysis. *International Journal of Health Geographics*, 10: 63.
7. Jiang, Z., Qiu, Z., Hatcher, J. (2010). Joinpoint Trend Analysis of Cancer Incidence And Mortality Using Alberta Data. *Cancer Surveillance, Surveillance and Health Status Assessment, Alberta Health Services*, 1-45.
8. Kim, H. J., Fay, M. P., Feuer, E. J., Midthune, D. N. (2000). Permutation tests for joinpoint regression with applications to cancer rates. *Statistics in Medicine*, 19: 335-351.
9. Kim, H. J., Fay, M. P., Yu, B., Barrett, M. J., Feuer, E. J. (2004). Comparability of Segmented Line Regression Models. *Biometrics*, 60(4): 1005-1014.
10. Marrett, L. D. (2010). Colorectal Cancer Network (CRCNet) User Documentation for Surveillance Analytic Software: Joinpoint. *Cancer Care Ontario*, 1-28.
11. Pickle L., Hao Y., Jemal A., Zou Z., Tiwari R.C., Ward E., Hachey M., Howe H.L., Feuer E.J. (2007). A New Method of Estimating United States And State-Level Cancer Incidence Counts for the Current Calendar Year CA: A. *Cancer Journal of Clinicians*, 57: 30-42.
12. Vural, A. (2007). Aykırı Değerlerin Regresyon Modellerine Etkileri ve Sağlık Kestiriciler, Yüksek Lisans Tezi, Marmara Üniversitesi, Sosyal Bilimler Enstitüsü, İstanbul.