

Fall 2015

# The Sensitivity of a Test Based on Spearman's Rho in Cross-Correlation Change Point Problems

Congjian Liu

Follow this and additional works at: <https://digitalcommons.georgiasouthern.edu/etd>



Part of the [Applied Statistics Commons](#)

---

## Recommended Citation

Liu, Congjian, "The Sensitivity of a Test Based on Spearman's Rho in Cross-Correlation Change Point Problems" (2015). *Electronic Theses and Dissertations*. 1336.  
<https://digitalcommons.georgiasouthern.edu/etd/1336>

This thesis (open access) is brought to you for free and open access by the Graduate Studies, Jack N. Averitt College of at Digital Commons@Georgia Southern. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons@Georgia Southern. For more information, please contact [digitalcommons@georgiasouthern.edu](mailto:digitalcommons@georgiasouthern.edu).

THE SENSITIVITY OF A TEST BASED ON SPEARMAN'S RHO IN  
CROSS-CORRELATION CHANGE POINT PROBLEMS

by

CONGJIAN LIU

(Under the Direction of Patricia Humphrey)

ABSTRACT

In change point problems, there are three main questions that researchers are interested in. First of all, is there a change point or not? Second, when does the change point occur in a time series? Third, how quickly can we detect the change point? In this thesis, we first explain what a change point is, and what a cross-correlation is. We then discuss prior research in this area. Then we discuss and examine a test based on Spearman's rho, introduced by Wied and Dehling (2011), which tests the null hypothesis of no change point, and compare the change point we set with the results from the proposed statistic in simulation. We also use this statistic on data we selected from the U.S. stock market. We conclude with the pros and cons of this statistical method, and how we can detect the change point sensitively using the proposed statistic.

KEY WORDS: CHANGE POINT, CROSS-CORRELATION, U.S. STOCK PRICE

THE SENSITIVITY OF A TEST BASED ON SPEARMAN'S RHO IN  
CROSS-CORRELATION CHANGE POINT PROBLEMS

by

CONGJIAN LIU

B.A., Sichuan University, P.R. China, 2007

A Thesis Submitted to the Graduate Faculty of Georgia Southern University in  
Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

STATESBORO, GEORGIA

©2015

CONGJIAN LIU

All Rights Reserved

THE SENSITIVITY OF A TEST BASED ON SPEARMAN'S RHO IN  
CROSS-CORRELATION CHANGE POINT PROBLEMS

by

CONGJIAN LIU

Major Professor: Patricia Humphrey  
Committee: Charles Champ  
Arpita Chatterjee

Electronic Version Approved:  
Fall 2015

## ACKNOWLEDGMENTS

First, I would like to thank Dr. Patricia Humphrey for making this thesis possible. It was only through her supervision and exhortation that this was possible. I would also like to thank my committee for their valuable advice.

## TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS.....	5
LIST OF TABLES .....	7
LIST OF FIGURES.....	8
CHAPTER	
1. Introduction .....	9
Background .....	9
Purpose of the study.....	17
2. Methodology.....	20
3. Simulation Study.....	24
Simulation procedure .....	24
Simulation results and analysis.....	26
4. Application to Real Data.....	35
5. Summary and Discussion.....	43
REFERENCES.....	45
APPENDIX A   Code for Simulation.....	48
APPENDIX B   Results of Simulations.....	60
APPENDIX C   Results of Different Time Intervals.....	65

## LIST OF TABLES

Table	Page
Table 3.1 Parameter Settings for Small Sample Size.....	26
Table 3.2 Results of Effects of $\gamma_n$ with $n_1 = 25, n_2 = 5, 25, 50,$ $r_1 = 0.5, r_2 = -0.5$ .....	27
Table 4.1 Cross-correlation for the Full Dataset Due to the Visual Change Point, Jan. 1 <sup>st</sup> , 2013.....	36
Table 4.2 Estimated Change Point for Different Time Intervals.....	37
Table 4.3 Cross-Correlation for the Reduced Dataset from September 4 <sup>th</sup> , 2012 to May 3 <sup>rd</sup> , 2013.....	39
Table 4.4 Cross-Correlation for the Reduced Dataset from Dec. 3 <sup>rd</sup> , 2012 to Jan. 31 <sup>st</sup> , 2013.....	40
Table 4.5 Cross-Correlation Analysis for Different Time Intervals around April 1 <sup>st</sup> , 2013.....	41
Table 4.6 Estimated Change Point for Different Time Intervals around April 1 <sup>st</sup> , 2013.....	42



## LIST OF FIGURES

Figures	Page
Fig. 1.1 A simple Example for Change Point of Continuous Variable.....	10
Fig. 1.2 A simple Example for Change Point of Discontinuous Variable.....	10
Fig. 1.3 Dataset y1.....	11
Fig. 1.4 Example plots of Correlation.....	14
Fig. 1.5 An Example plot of Cross-Correlation.....	15
Fig. 1.6 Daily market price of Johnson & Johnson(JNJ) and Eli Lilly and Company (LLY) from 2009.4 to 2014.4.....	18
Fig. 4.1 Plot of LLY and JNJ Stock Prices From September 4 <sup>th</sup> , 2012 to May 3 <sup>rd</sup> , 2013.....	38
Fig. 4.2 Plot of LLY and JNJ Stock Prices From Mar. 1 <sup>st</sup> , 2013 to Jul. 31 <sup>st</sup> , 2013.....	41

## CHAPTER 1

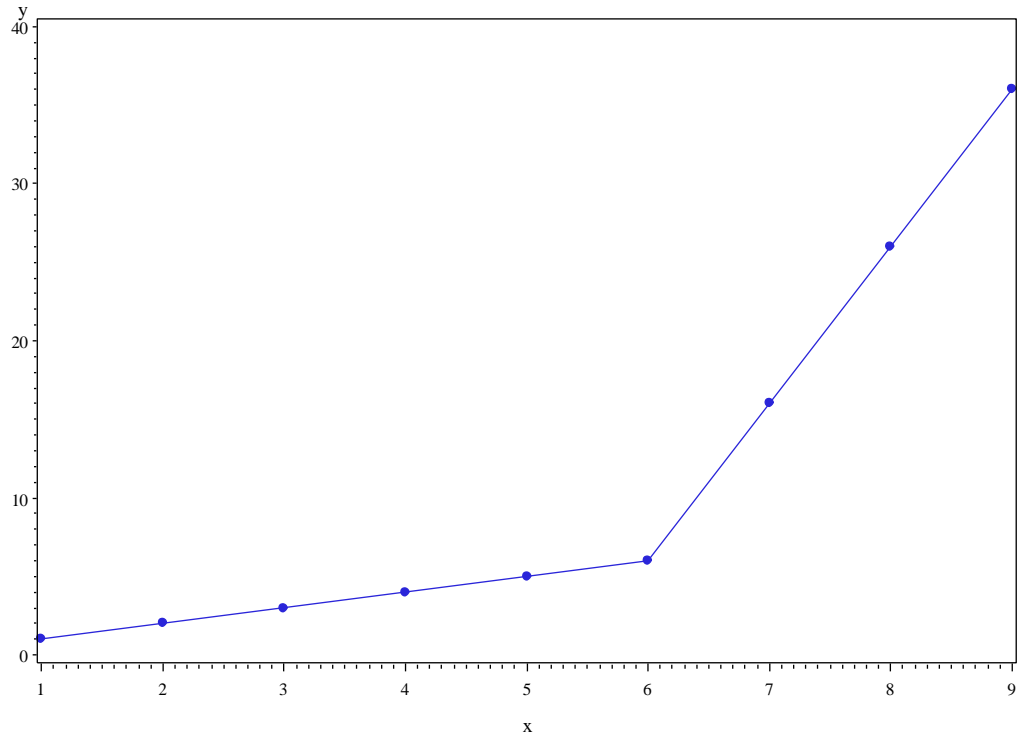
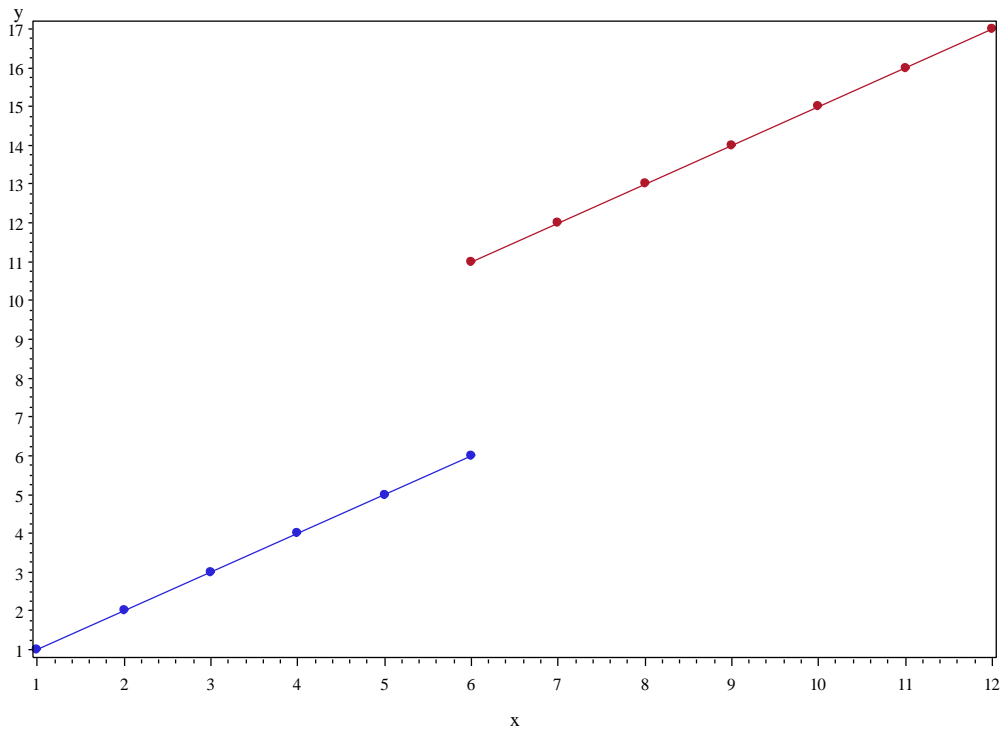
### INTRODUCTION

#### 1. Background

##### 1.1 Change point

Change point problems are problems with data collected over a period of time or space when there is a possible change in the underlying process/ population. It is of interest to analyze inferences about the point in the sequence where the change occurs (Everitt and Skrondal 2010). Research on change point problems was initially motivated by quality control in the industrial sector. Applications of change points occur in many areas such as: economics, finance, medicine, biology and computer science. Results of change point research also applies to the stock market. This helps with the observation of fluctuating price relative to time.

Hofrichter (2007) pointed out that change point models in a 2-dimensional dataset can be divided into two groups. First, models with a continuous change at the change point (Fig. 1.1), and second, with discontinuous change at the change point (Fig. 1.2). In both of the following plots, it is clearly shown that when  $x = 6$ , there is a change point.

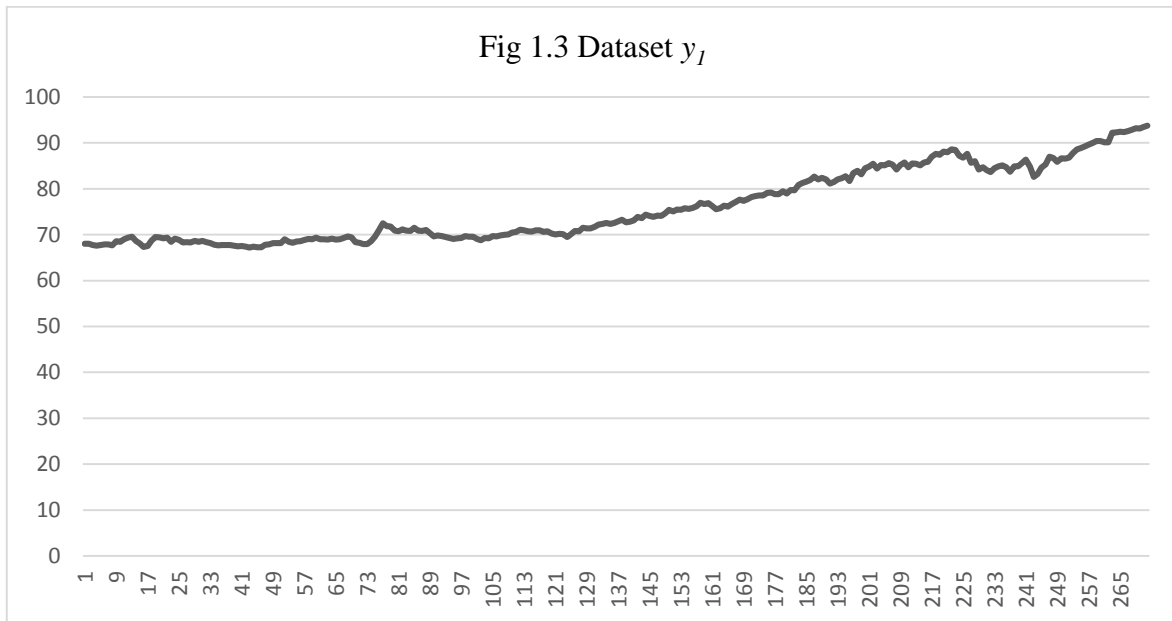
**Fig. 1.1 A Simple Example of a Continuous Change Point****Fig. 1.2 A Simple Example of a Discontinuous Change Point**

Mathematically, a change point implies a location or time. Data follow two different models, one model before the change point and a different one after the change point. For instance, consider the following linear regression model. Assume the model for a data set is the following:

$$\begin{cases} y_i = \alpha_1 + \beta_1 x_i + \varepsilon_{1i}, & i = 1, 2, \dots, t \\ y_i = \alpha_2 + \beta_2 x_i + \varepsilon_{2i}, & i = t + 1, t + 2, \dots, T \end{cases}$$

In the above model,  $\alpha$ 's and  $\beta$ 's are model parameters assumed to be fixed,  $\varepsilon$ 's are noise. The point  $(x_t, y_t)$  is the change point.

However, in general, detecting a change point of actual data is not always so easy. For example, if we have a plot of the dataset  $y_I$  in Fig. 1.3,



we can not definitely detect the main structural change point of the dataset. Actual data is like this. The only thing we can point out is that there may be a main structural change point somewhere around  $x = 6$ . However, it also may be the case that the correct population model

is nonlinear and there is no change point. Hence, we want to use some statistical methods to test for a significant change point.

Along with detecting that there is a change in a process, researchers also want to know the exact time when the change happened. Pettitt (1979) introduced a nonparametric method based on a modification of the Mann-Whitney two-sample test in the change point problem. In this test we set two samples  $X_1, \dots, X_t$  and  $X_{t+1}, \dots, X_T$  ( $1 \leq t < T$ ) that come from the same population. Let:

$$D_{ij} = \text{sgn}(x_i - x_j)$$

where

$$\text{sgn}(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{if } x = 0 \\ -1, & \text{if } x < 0 \end{cases}$$

and consider

$$U_{t,T} = \sum_{i=1}^t \sum_{j=t+1}^T D_{ij}.$$

Then Pettitt (1979) proposes the test of  $H_0$ : no change vs.  $H_A$ : change by using the statistic

$$K_T = \max_{1 \leq t < T} |U_{t,T}|$$

Pettitt (1979) also considered the condition for Bernoulli and Binomial observations. Since in the quality control area, manufacturer cares whether products are good or bad when they come out from the assembly line, we want to use this nonparametric method for Binomial observations, because we consider the parameters of Binomial are  $n$  and  $p$ , which are the number of trials and success probability in each trial. We can use this method in the quality control area to detect a change in  $p$ . This is a major consideration for the manufacturer in actual industrial production. Daniel Barry and J. A. Hartigan proposed a Bayesian analysis for change

point problems, especially in detecting sharp short-lived changes in the parameters (Hartigan, 1993). Nobuyuki Tamura et al. showed how to use a Bayesian information criterion to solve the change point problem for failure-censored data which followed an exponential distribution. (Tamura, 2008). Another way to study change point problems is discrete wavelet analysis theory. Maitreyi Sur et al. proposed that discrete wavelet analysis could be used in change point problems in the case of animals moving through a complex environment (Maitreyi Sur, 2014).

In our following study, we evaluated the ability of a test based on Spearman's rho proposed by Wied et al, (2011) to detect the change point in simulation and real data.

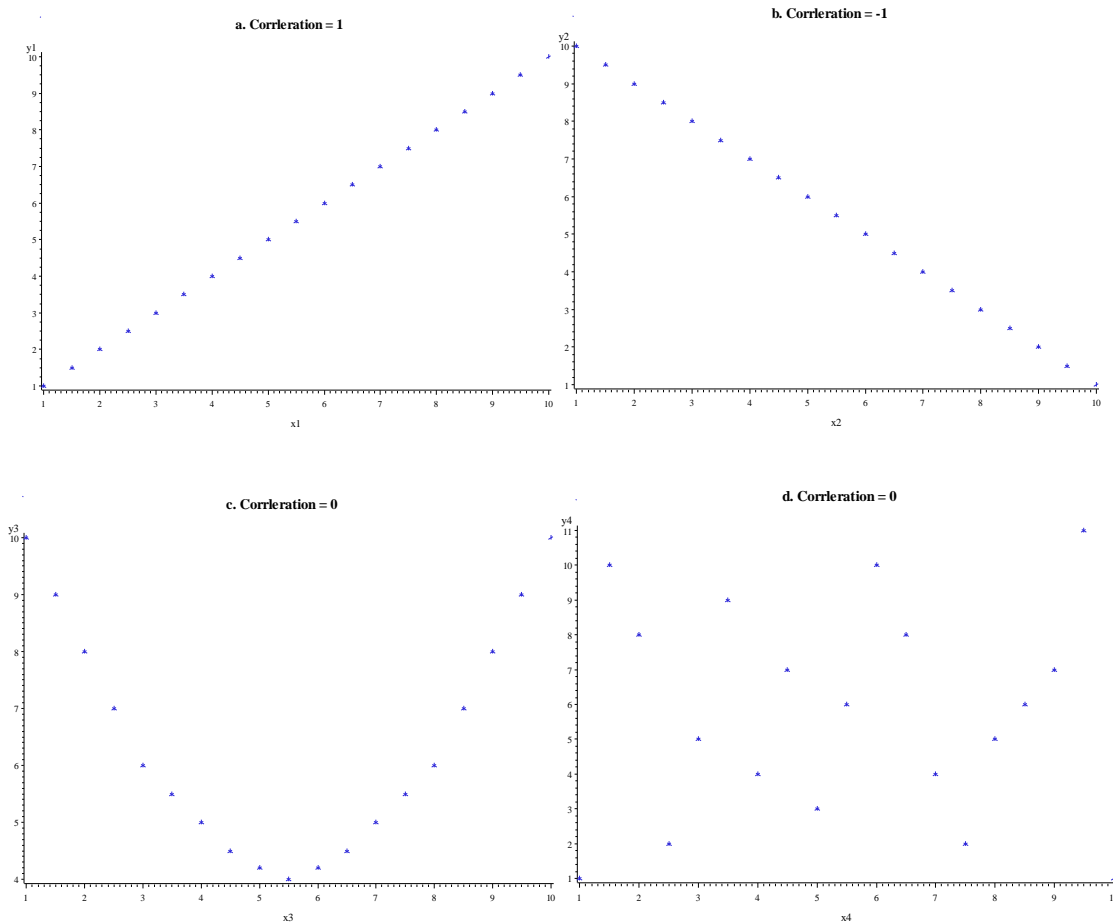
## 1.2 Cross-correlation in the change point problem

Cross-correlation is a statistical measurement to compare and relate one time series of data with another time series of data with possible time lag. It was first introduced in the 1960s by Papoulis (1962) and Bracewell (1965). Correlation is the degree to which two measurements in the same group of elements show a tendency to vary together linearly. The Pearson product-moment correlation coefficient is defined as (Rodgers, 1988) :

$$\rho = \frac{E(X_i - \mu_x)(Y_i - \mu_y)}{[E(X_i - \mu_x)^2 E(Y_i - \mu_y)^2]^{1/2}}$$

where  $X_i$  and  $Y_i$  are two datasets for  $i = 1, 2 \dots n$ . Here we draw correlation plots for example, see Fig. 1.4.

Fig 1.4 Example Plots of Correlation



The classical Pearson's correlation coefficient is between -1 and 1. When the coefficient equals 0, you do not know whether there is no relationship as in Fig. 1.4d or a nonlinear relationship as in Fig. 1.4c. Nonlinear means that there is no linear relationship between  $x$  and  $y$  datasets. When  $\rho = 1$ , there is a perfect positive linear relationship and when  $\rho = -1$ , there is a perfect negative linear relationship. Hence, the correlation coefficient provides us a quantification of the strength of the linear relationship between one variable and another one.

The cross-correlation coefficient is usually defined as (Tsay, 2010):

$$\rho(\tau) = \frac{E[(X_t - \mu_X)(Y_{t+\tau} - \mu_Y)]}{(\sigma_X \sigma_Y)}$$

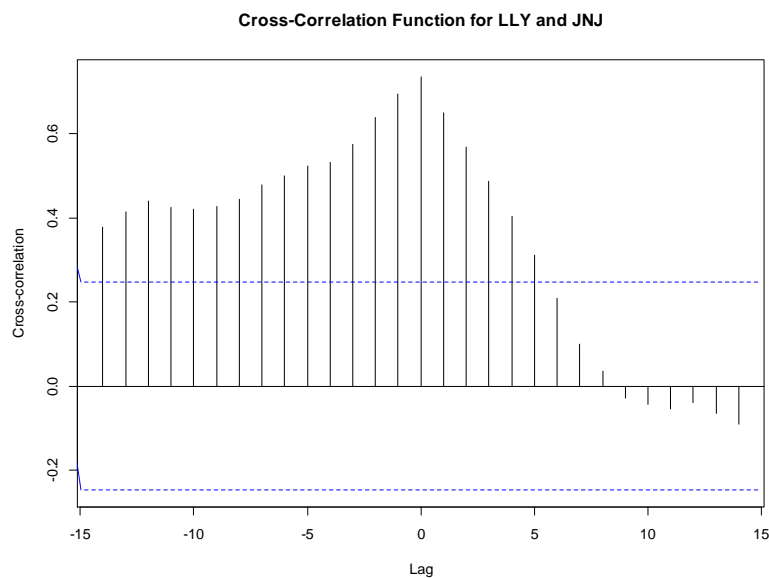
where  $X_t$  and  $Y_t$  are two time series datasets with integer time index  $t$  and the time lag parameter  $\tau$ ,  $\sigma$  is the standard deviation of each dataset. When  $\tau$  is positive,  $X_t$  are predictors of  $Y_t$ , it is sometimes said that  $X$  leads  $Y$ , and with  $\tau$  negative, we said that  $Y$  leads  $X$ .

From the definitions above, we can conclude that cross-correlation is similar to regular correlation except for time lag effects.

We note that the cross-correlation coefficient (CCF), is between -1 and 1, as the regular correlation. The meaning of the cross-correlation coefficient is also similar to the correlation coefficient except we must point out the time lag when we interpret the cross-correlation coefficient. Hence, the cross-correlation coefficient provides us with a quantification of the level of relationship between two time series of data with possible time lag.

Figure 1.5 depicts an example of cross-correlation. The  $X$  dataset in Fig. 1.5 is Eli Lilly and Company (LLY)'s stock price and  $Y$  dataset is Johnson & Johnson (JNJ)'s stock price from April 1<sup>st</sup>, 2009 to June 30<sup>th</sup>, 2009.

Fig. 1.5 An Example plot of cross-correlation





From the plot above, when lag = 0, the cross-correlation coefficient (CCF) is 0.7359, which means Eli Lilly and Company and Johnson & Johnson's stock prices change in the same direction at the same time. This may be due to the changes of the whole stock market index.

Cross-correlation analysis in financial markets has been used in many studies. Li Liu studied cross-correlations between crude oil and agricultural commodity markets (Liu, 2014). She used Podobnik's Q test which can be easily implemented to detect linear cross-correlations between crude oil and agricultural commodities, then used detrended cross-correlation analysis (DCCA) proposed by Podobnik and Stanley to analyze those markets. Data in her paper were daily closing spot prices of crude oil and four agricultural commodities taking corn, soybean, oat and wheat as examples from January 3, 1994 to December 31, 2012. The results showed that at smaller lag length, the linear cross-correlations are not significant, while at large lag lengths, they are significant. The volatility cross-correlations are highly significant at all lag lengths. In addition, nonlinear cross-correlations are weak but significant for small time intervals from 4 to 128 days and not significant for large time intervals (about one year). Liu pointed out that their analysis supports studies which suggest that the food crisis was caused by high crude oil prices. Marwane El Alaoui and Saâd Benbachir mainly investigated cross-correlation among four financial markets from the MENA (Middle East and North Africa) area via Multifractal detrended cross-correlation analysis (MF-DCCA) (El Alaoui, 2013). MF-DCCA between the four markets shows these countries' markets have small but significant cross-correlations. The cross-correlation coefficient is very close to zero ( $\rho < 0.05$ ) for Amman Stock Exchange (AMX) and Tunis Stock Exchange (TUNINDEX), Casablanca Stock Exchange (MASI) and Tunis Stock Exchange (TUNINDEX) and Tunis Stock Exchange (AMX) and Casablanca Stock Exchange (MASI). But for other pairs, like AMX and the

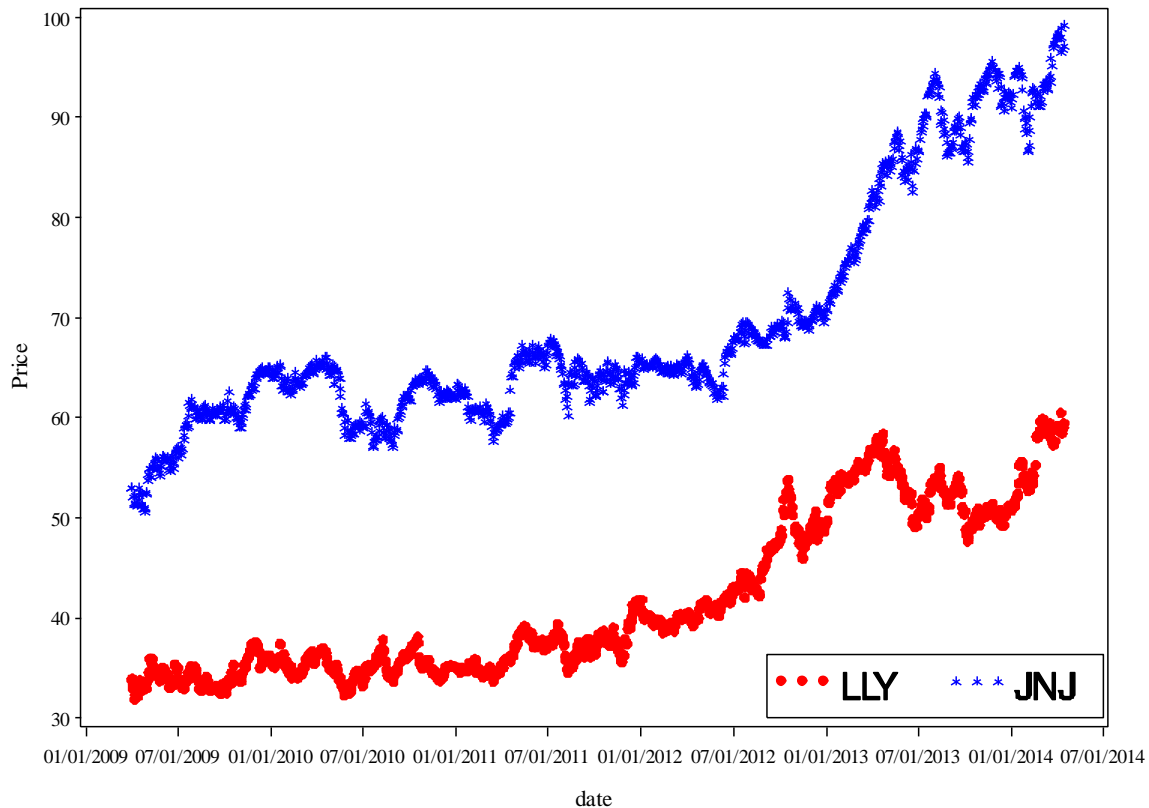
Egyptian Exchange (EGX), MASI and EGX, and TUNINDEX and EGX, the cross-correlations exceed 0.15 when time lag is greater than 800. The analysis reaches a coherent conclusion with the fact that, being the four signatories of a free trade agreement, the bonds between their markets are weak and their markets need further development. In sum, that paper shows that, in the case of MENA markets, the MF-DCCA method is valid. Boris Podobnik et al. did cross-correlation analysis between volume growth rate  $\tilde{R}$ , which is the difference in logarithm between two consecutive values of trading volume, and price change  $R$  in financial market (Podobnik et al., 2009). Podobnik et al. assumed that the underlying processes for logarithmic price change  $|R|$  and logarithmic volume change  $|\tilde{R}|$  are similar, and used a new method based on DCCA, which applied the cross-correlation analysis in the financial market as a whole, not for individual companies. By analyzing the Standard and Poor's (S & P) 500 Index over the 59-year period 1950–2009, power-law cross-correlations between properties of volume growth rate  $|\tilde{R}|$  and properties of price change  $|R|$  are found, while no valid relation is found between  $\tilde{R}$  and  $R$ . In addition, the cross-correlations between  $|R_{t+n}|$  and  $|\tilde{R}_t|$  are not only significant at zero time lag ( $n = 0$ ) but for a large range of time lags, where  $n$  is time lag around time  $t$ . Yet the author failed to reach a coherent conclusion for the central and tail part of the probability density function  $P(|\tilde{R}|)$ .

## 2. Purpose of the Study

In our application of interest, we want to find the main structural change point between two stocks' prices with minimum observations after the true change point. We hope to achieve our objective though analyzing the cross-correlation before and after a change point. Recent five-year daily stock prices of Johnson and Johnson (JNJ) and Eli Lilly and Co. (LLY) in U.S. NASDAQ stock exchange from April 1<sup>st</sup>, 2009 to April 15<sup>th</sup>, 2014 were chosen as our study

subjects. It would be the first time to study these two stock prices via cross-correlation. These two companies are both well-known companies in pharmaceuticals. Their products share similar features: for instance, some of their medicine products target at the same kind of consumers in the market. It is of interest to study how these two stocks affect each other in terms of change points and cross-correlation, and how the other one reacts when one of these stocks decreases. The daily price for the last five years of JNJ and LLY are shown in Figure 1.6. Overall, their prices have been increasing over the last 5 years.

**Fig. 1.6 Daily market price of Johnson & Johnson (JNJ) and Eli Lilly and Company (LLY) from 2009.4 to 2014.4**



We focus on the main structural change point which appears to occur about Jan.1<sup>st</sup> 2013, because before Jan.1<sup>st</sup> 2013, JNJ and LLY move relatively close, but JNJ sharply rises and LLY has a decrease after Jan.1<sup>st</sup> 2013.

In this paper, we will investigate the behavior of a test, which is based on Spearman's rho to test the change point with simulation. We want to know how the number of observations before and after the change point affects detecting a known change point; also we would like to know how varying the difference in cross-correlation before and after a known change point affects detection. Then we apply this method to the study of Johnson & Johnson (JNJ) and Eli Lilly and Co. (LLY). The historical data for these stock prices can be found and downloaded from the NASDAQ website (<http://www.nasdaq.com/>).

The remainder of this paper is organized as follows: Chapter 2 is a description of the methodology. It provides mathematical functions to describe change points, cross-correlation and the testing method to determine change points. Chapter 3 is our simulation results. The simulation includes using the proposed test to investigate sensitivity to both the number of data values and change in the cross-correlation. Chapter 4 uses the method on JNJ and LLY's stock prices. Chapter 5 is a summary of this study and future research interests.

## CHAPTER 2

### METHODOLOGY

A test based on Spearman's rho is used to determine the change point as proposed by Wied and Dehling. (Dominik Wied et al, 2011). They proposed a cumulative sum (CUSUM) method. Spearman's rho is probably the most common rank-based measure of association in economic and social sciences, as it has a much better performance in the presence of outliers and it has no other condition on the existence of moments.

Suppose we select samples from underlying population, denoted as  $x_n$ ,  $n = 1, 2, \dots, N$ . To detect a change point, partition the data into  $x_1, x_2, \dots, x_k$  and  $x_{k+1}, x_{k+2}, \dots, x_n$  for  $k = 2, 3, \dots, N-2$ .

We want to test if there is a change point between time  $k$  and  $k+1$ . For easy interpretation, we define the change point at time  $k$ . Let  $d$  denote the dimension of the data in the sample. For instance, with LLY and JNJ, we have two stocks, so  $d = 2$ .

Define In terms of the rank, Spearman's rho up time  $k$  is defined as

$$\rho_k = h(d) * \left( \frac{2^d}{k} \sum_{j=1}^k \prod_{i=1}^d (1 - \hat{U}_{i,j;n}) - 1 \right)$$

with

$$h(d) = \frac{d + 1}{2^d - (d + 1)}$$

and

$$\hat{U}_{i,j;n} = \frac{1}{n} (\text{rank of } X_{i,j} \text{ in } X_{i,1} \dots X_{i,n}), i = 1, \dots, d, j = 1, \dots, n$$

where  $k$  is the  $k^{\text{th}}$  observation,  $d$  is the dimension of the data, and  $n$  is the total number of observations.  $i$  and  $j$  are ranking indices. Rank in a dataset means the rank of the smallest observation is 1; the rank of largest one is  $n$ .

Now, we want to test whether Spearman's rho is significantly different in two segments.

The null hypothesis is:

$$H_0: \rho_k = \rho_0, k = 1, \dots, n$$

The alternative hypothesis is

$$H_a: \exists k \in \{1, 2 \dots n - 1\}, \rho_k \neq \rho_{k+1}$$

If the null hypothesis is rejected, there is a change in the cross-correlation, and the  $k^{\text{th}}$  observation is where the change occurs. If the null hypothesis fails to be rejected, there is no change in cross-correlation.

Wied and Dehling proposed the statistic test  $W$  as:

$$W = \hat{D} \max_{1 \leq k \leq n} \left| \frac{k}{\sqrt{n}} (\hat{\rho}_k - \hat{\rho}_n) \right|;$$

With the value of  $\hat{D}$  is defined as:

$$\hat{D} = \frac{1}{\sqrt{\hat{D}'}}$$

If  $W = 0$  we fail to reject the null hypothesis. If  $W \neq 0$ , we reject the null hypothesis, and the observation  $k$  is a change point in cross-correlation.

where

$$\begin{aligned} \widehat{D}' = h(d)^2 2^{2d} & \left\{ \frac{1}{n} \sum_{j=1}^n \prod_{i=1}^d (1 - \widehat{U}_{i,j;n})^2 - \left( \frac{1}{n} \sum_{j=1}^n \prod_{i=1}^d (1 - \widehat{U}_{i,j;n}) \right)^2 \right. \\ & + 2 \left[ \sum_{m=1}^{\gamma_n} k\left(\frac{m}{\gamma_n}\right) \left( \sum_{j=1}^{n-m} \frac{1}{n} \prod_{i=1}^d (1 - \widehat{U}_{i,j;n}) (1 - \widehat{U}_{i,j+m;n}) \right. \right. \\ & \left. \left. - \left( \frac{1}{n} \sum_{j=1}^n \prod_{i=1}^d (1 - \widehat{U}_{i,j;n}) \right)^2 \right) \right] \left. \right\} \end{aligned}$$

Where the  $\gamma_n$  is chosen such that  $\gamma_n = o(n^{1/2})$ , from the definition of function  $o(\bullet)$ ,  $\gamma_n \leq \sqrt{(n)}$  where  $n = n_1 + n_2$ . And the kernel function  $k(\bullet)$  is contained in the class  $\mathcal{K}_2$  of Andrews (1991), which includes the following:

Truncated:  $k(y) = \begin{cases} 1, & \text{for } |y| \leq 1 \\ 0, & \text{otherwise} \end{cases}$

Bartlett:  $k(y) = \begin{cases} 1 - |y|, & \text{for } |y| \leq 1 \\ 0, & \text{otherwise} \end{cases}$

Parzen:  $k(y) = \begin{cases} 1 - 6y^2 + 6|y|^3, & \text{for } 0 \leq |y| \leq 1/2 \\ 2(1 - |y|)^3, & \text{for } 1/2 \leq |y| \leq 1 \\ 0, & \text{otherwise} \end{cases}$

Tukey-hanning:  $k(y) = \begin{cases} \frac{1 + \cos(\pi y)}{2}, & \text{for } |y| \leq 1 \\ 0, & \text{otherwise} \end{cases}$

Quadratic Spectral:  $k(y) = \frac{25}{12\pi^2 y^2} \left( \frac{\sin(6\pi y/5)}{6\pi y/5} - \cos\left(\frac{6\pi y}{5}\right) \right)$

In a general dataset, we set  $x_t$  as the value of the variable in period  $t$ , and  $x_{t-p}$  is the value of the variable in period  $t-p$  or “lagged  $p$  period” or “lagged X.” The AR( $p$ ) model of this general dataset is:

$$x_t = \alpha + \beta x_{t-p} + \varepsilon$$

Most stock data is “best” modeled as AR (1) time series, which means time lag  $p = 1$ . In our study, the model of this condition is:

$$x_t = \alpha + \beta x_{t-1} + \varepsilon$$

Since our study is based on lag truncation parameter  $p = 1$  and for simplicity, we choose the truncated kernel function in our study.



## CHAPTER 3

### SIMULATION STUDY

#### 3.1 Simulation procedure

We constructed simulations using FORTRAN (version: FTN 95), the Plato IDE compiler and the following procedure:

1. Generate two sets of sample data with a given change point. We define  $n_1$  as the number of observations before the change point;  $n_2$  is number of observations after the change point. The cross-correlation between the two sets before the change point is  $r_1$ ;  $r_2$  is the cross-correlation between the two sets after the change point. The bandwidth  $\gamma_n$  is chosen such that  $\gamma_n = o(n^{1/2})$ , so  $2 \leq \gamma_n \leq \sqrt{(n_1 + n_2)}$ .
2. The simulation data were created using the following function:

$$yy1 = \sqrt{(-2.0 * \log(e)) * \cos(2 * e1 * \pi)} \quad \text{and}$$

$$yy2 = \sqrt{(-2.0 * \log(e)) * \cos(2 * e1 * \pi)}$$

with  $e$  and  $e1$  as random numbers which were created by the FORTRAN system with Windows operating system clock time seed. This means that  $yy1$  and  $yy2$  are independent and follow a normal distribution  $N(0, 1)$ .

Now we set  $x(i) = yy1$  and create  $y(i)$  as a function of  $x(i)$  with given cross-correlation coefficients  $r_1$  and  $r_2$ . When processing step  $i \leq n_1$ ,

$$y(i) = x(i) * r_1 + yy2 * c_1 \quad \text{otherwise,}$$

$$y(i) = x(i) * r_2 + yy2 * c_2$$

where  $c_1$  and  $c_2$  are coefficients,  $c_i = \sqrt{1 - r_i^2}$ . Here,  $x(i)$  and  $y(i)$  are two sets of data.

3. To compute the statistic test  $W$  for that simulation sample, the subroutines rank and sort were downloaded from the National Institute of Standards and Technology website (<http://www.nist.gov/>).

Select different values of  $n_1$  and  $n_2$ ,  $r_1$  and  $r_2$  to construct different datasets to test the change point. The following table shows the conditions we used in the simulation. We didn't set  $n_2$  greater than 50, because in the real world, 50 business days is more than two months after a true change point. We want to detect the change point as soon as possible to make a profit or to save funds. So it can be concluded that a change point detected after 50 business days is useless; ergo,  $n_2 \leq 50$ . Table 3.1 shows the parameter settings for small sample sizes.

Since we want to analyze the stock prices in the U.S. stock market, and almost every stock price has a large amount of historical data, we also set  $n_1 = 200, 500$ ,  $n_2 = 5, 10, 20, 50$ ;  $r_1$  and  $r_2$  are the same as above.

4. Repeat above steps for each scenario 1000 times.
5. Use SAS 9.3 to analyze the mean, standard deviation, the 25<sup>th</sup> percentile, 50<sup>th</sup> percentile and 75<sup>th</sup> percentile of the 1000 estimated change points.

Table 3.1 Parameter Settings for Small Sample Sizes

$n_1$	$n_2$	$r_1$	$r_2$	$n_1$	$n_2$	$r_1$	$r_2$
5	5	1	-1	5	5	0.5	-0.5
	25				25		
	50				50		
25	5	1	-1	25	5	0.5	-0.5
	25				25		
	50				50		
50	5	1	-1	50	5	0.5	-0.5
	25				25		
	50				50		
$n_1$	$n_2$	$r_1$	$r_2$	$n_1$	$n_2$	$r_1$	$r_2$
5	5	-0.7	0.4	5	5	0.4	0.7
	25				25		
	50				50		
25	5	-0.7	0.4	25	5	0.4	0.7
	25				25		
	50				50		
50	5	-0.7	0.4	50	5	0.4	0.7
	25				25		
	50				50		
$n_1$	$n_2$	$r_1$	$r_2$	$n_1$	$n_2$	$r_1$	$r_2$
5	5	0.2	0.8	5	5	-0.8	0.4
	25				25		
	50				50		
25	5	0.2	0.8	25	5	-0.8	0.4
	25				25		
	50				50		
50	5	0.2	0.8	50	5	-0.8	0.4
	25				25		
	50				50		

### 3.2 Simulation results and analysis

In this part, first we must define the word “catch” in the following, we got means ( $\bar{x}$ ) and standard deviations ( $s$ ) from the SAS output. When the true change point located in the interval  $(\bar{x} - s, \bar{x} + s)$ , we said we catch the true change point. The words “meaningful results” have the same implication as above.

### 3.2.1 $\gamma_n$ effects

Table 3.2 Results of Effects of  $\gamma_n$  with  $n_1 = 25$ ,  $n_2 = 5, 25, 50$ ,  $r_1 = 0.5$ ,  $r_2 = -0.5$

$n_1$	$n_2$	$\gamma_n$	Mean	Std. Dev.	25 <sup>th</sup> Pctl.	50 <sup>th</sup> Pctl.	75 <sup>th</sup> Pctl.
25	5	2	14.49	5.98	9	15	19
		3	14.62	6.15	10	15	20
		4	14.4	6.09	10	15	19
		5	14.46	6	10	15	19
	25	2	21.09	8.22	16	22	25
		3	21.77	7.95	17	22	26
		4	21.22	7.88	16	22	25
		5	21.63	8.36	16	22	26
		6	21.54	8.18	16	22	26
		7	21.09	8.34	16	21	25
	50	2	28.15	12.83	20	25	34.5
		3	28.21	12.47	20	25	35
		4	28.06	13.21	20	24	34
		5	28.05	13.03	20	24	34
		5	27.13	12.4	20	24	33
		7	28.21	12.75	20	25	35
		8	28.13	13.29	20	24	34

First of all, Table 3.2 shows the results of parameter settings in a small sample size with  $n_1 = 25$ ,  $n_2 = 5, 25, 50$ ,  $r_1 = 0.5$ ,  $r_2 = -0.5$  for example. We found a very interesting result:  $\gamma_n$  does not seem to affect the results of the test. This phenomenon appeared in all of the parameter settings we examined. Hence we concluded that  $\gamma_n$  does not affect the results of the test, and any reasonable  $\gamma_n$  can be used. We choose  $\gamma_n = 3$  in our following work. The tables of all results when  $\gamma_n = 3$  are in Appendix B.

### 3.2.2 The number of observations effect

From Appendix B, we found that the number of observations before the true change point,  $n_1$ , and the number of observations after the true change point,  $n_2$ , clearly affects the detection of the change point. When  $n_2$  is equal to or greater than half of  $n_1$  and less than or equal to

$n_1$ ;  $\frac{1}{2}n_1 \leq n_2 \leq n_1$ , meaningful results can be found. Appendix B shows the proposed method can catch the true change point under almost every  $r_1, r_2$  setting we examined, except when  $n_1 = n_2 = 5$ ,  $r_1 = 1$  and  $r_2 = -1$ . That may due to the very small sample size. Overall, this method effectively catches the true change point, especially when the number of observations after the change point is equal to the number of observations before the change point, which is  $n_1 = n_2$ .

### 3.2.3 The cross-correlation effect

In this section, we examined several conditions of  $r_1$  and  $r_2$  with varying signs and magnitudes of difference. The full results are given in the Appendix B. We discuss here some choices of  $n_1$  and  $n_2$  that illustrate our conclusion about the sample size given above. In “small sample size” cases, we will discuss  $n_1 > n_2$  with  $n_1 = 50, n_2 = 25$ ,  $n_1 = n_2$  with  $n_1 = n_2 = 25$ , and  $n_1 < n_2$  with  $n_1 = 5, n_2 = 50$  as examples. In “large sample size” cases, we will discuss  $n_1 = 500, n_2 = 20$  as an example.

#### 3.2.3.1 $r_1 = 1, r_2 = -1$

In this parameter setting, the difference between  $r_1$  and  $r_2$  is equal to 2, which is the largest possible difference. The cross-correlation before the true change point is the strongest positive, and the cross-correlation after the true change point is the strongest negative.

When  $n_1 = 50, n_2 = 25$ , the mean of the estimated change point is 42.88 with standard deviation 7.35. The median of the estimated change point is 46. This simulation caught the true change point which is  $k = 50$ . The condition of the number of observation in this result satisfies  $\frac{1}{2}n_1 \leq n_2 \leq n_1$ .

For  $n_1 = n_2 = 25$ , the mean of the estimated change point is 21.24 with standard deviation 3.91. The median of the estimated change point is 23. This simulation caught the true change point which is  $k = 25$ . The condition of the number of observation in this result satisfies  $n_1 = n_2$ .

For  $n_1 < n_2$ , we use  $n_1 = 5, n_2 = 50$  as an example. The mean of the estimated change point is 12.89 with standard deviation 11.82. The median of the estimated change point is 7.5. This simulation caught the true change point which is  $k = 5$ . The condition of the number of observation in this result did not satisfies  $\frac{1}{2}n_1 \leq n_2 \leq n_1$ , and the large difference between  $n_1$  and  $n_2$  caused this biased result.

For  $n_1 > n_2$  with a large sample size. The mean of the estimated change point at this parameter setting is 329.38 with standard deviation 123.43. The median of the estimated change point is 350. This simulation does not catch the true change point which is  $k = 500$ . This result may due to the large difference between  $n_1$  and  $n_2$ .

### 3.2.3.2 $r_1 = 0.5, r_2 = -0.5$

In this parameter setting, the difference between  $r_1$  and  $r_2$  is equal to 1, which is half of the largest possible difference between  $r_1$  and  $r_2$ , with the same signs as given in 3.2.3.1.

For  $n_1 > n_2$  with  $n_1 = 50, n_2 = 25$ , the mean of the estimated change point is 37.95 with a standard deviation 12.31. The median of the estimated change point is 41. This simulation caught the true change point which is  $k = 50$ . The condition of the number of observations in this result satisfies  $\frac{1}{2}n_1 \leq n_2 \leq n_1$ , but the accuracy and sensitivity of this test is comparatively less than that of the result of  $r_1 = 1, r_2 = -1$  setting.

For  $n_1 = n_2 = 25$ , the mean of the estimated change point is 21.77 with standard deviation 7.95. The median of the estimated change point is 22. This simulation caught the true change point which is  $k = 25$ .

For  $n_1 < n_2$ , we also used  $n_1 = 5, n_2 = 50$  as an example. The mean of the estimated change point is 23.97 with standard deviation 12.89. The median of the estimated change point is 24. This simulation does not catch the true change point which is  $k = 5$ . We did not get meaningful results because the condition of the number of observation in this result did not satisfies  $\frac{1}{2}n_1 \leq n_2 \leq n_1$ , and the large difference between  $n_1$  and  $n_2$  might cause this biased result.

For large sample sizes, the mean of the estimated change point is 275.7 with standard deviation 117.52. The median of the estimated change point is 278.5. This simulation does not catch the true change point which is  $k = 500$ . This result may due to the large difference between  $n_1$  and  $n_2$ , that does not satisfies the  $\frac{1}{2}n_1 \leq n_2 \leq n_1$ .

### 3.2.3.3 $r_1 = -0.7, r_2 = 0.4$

In this parameter setting, the cross-correlation before the true change point is strong negative, and the cross-correlation after the true change point is positive.

For  $n_1 > n_2$ , the mean of the estimated change point is 47.71 with standard deviation 10.86. The median of the estimated change point is 49. This simulation caught the true change point which is  $k = 50$ . The condition of the number of observation in this result satisfies  $\frac{1}{2}n_1 \leq n_2 \leq n_1$ , and the accuracy and sensitivity of the test here is better than the result of  $r_1 = 1, r_2 = -1$  setting.

For  $n_1 = n_2$ , the mean of the estimated change point is 28.09 with standard deviation 7.35. The median of the estimated change point is 27. This simulation caught the true change point which is  $k = 25$ .

For  $n_1 < n_2$  with  $n_1 = 5$ ,  $n_2 = 50$ , the mean of the estimated change point is 28.27 with standard deviation 11.85. The median of the estimated change point is 29. This simulation does not catch the true change point which is  $k = 5$ . We did not get the meaningful results because the condition of the number of observation in this result does not satisfy  $\frac{1}{2}n_1 \leq n_2 \leq n_1$ , and the large difference between  $n_1$  and  $n_2$  might cause this biased result.

For large sample sizes, the mean of the estimated change point is 349.54 with standard deviation 125.64. The median of the estimated change point is 374.5. This simulation does not catch the true change point which is  $k = 500$ . This result due to the large difference between  $n_1$  and  $n_2$ , that does not satisfy the  $\frac{1}{2}n_1 \leq n_2 \leq n_1$ .

#### 3.2.3.4 $r_1 = 0.4, r_2 = 0.7$

In this parameter setting, the  $r_1$  and  $r_2$  are very close to each other and have same sign.

When  $n_1 > n_2$ , the mean of the estimated change point is 38.86 with standard deviation 16.51. The median of the estimated change point is 40. This simulation caught the true change point which is  $k = 50$ , but the accuracy and sensitivity of the test here is not as good as the results of previous parameter settings.

For example with  $n_1 = n_2 = 25$ , the mean of the estimated change point is 26.02 with standard deviation 10.54. The median of the estimated change point is 26. This simulation caught the true change point which is  $k = 25$ .



For  $n_1 < n_2$  with  $n_1 = 5$ ,  $n_2 = 50$ , the mean of the estimated change point is 27.96 with standard deviation 11.76. The median of the estimated change point is 28. This simulation does not catch the true change point which is  $k = 5$ . We did not get the meaningful results because of the same reason as other  $n_1 < n_2$  examples.

For  $n_1 = 500$ ,  $n_2 = 20$ , the mean of the estimated change point is 275.5 with standard deviation 117.52. The median of the estimated change point is 278.5. This simulation does not catch the true change point which is  $k = 500$ . This result might due to the large difference between  $n_1$  and  $n_2$ , that does not satisfy the  $\frac{1}{2}n_1 \leq n_2 \leq n_1$ .

### 3.2.3.5 $r_1 = 0.2, r_2 = 0.8$

In this parameter setting, the  $r_1$  and  $r_2$  have a big difference, but both of them have the same sign.

For  $n_1 > n_2$ ,  $n_1 = 50$ ,  $n_2 = 25$ , the mean of the estimated change point is 47.71 with standard deviation 10.86. The median of the estimated change point is 49. This simulation caught the true change point which is  $k = 50$ .

We again use  $n_1 = n_2 = 25$ , the mean of the estimated change point is 28.09 with standard deviation 7.35. The median of the estimated change point is 27. This simulation caught the true change point which is  $k = 25$ .

For  $n_1 < n_2$  with  $n_1 = 5$ ,  $n_2 = 50$ , the mean of the estimated change point is 28.27 with standard deviation 11.85. The median of the estimated change point is 29. This simulation does not catch the true change point which is  $k = 5$ . We did not get the meaningful results because of the same reason as other  $n_1 < n_2$  examples.

For large sample sizes, the mean of the estimated change point is 349.54 with standard deviation 125.64. The median of the estimated change point is 374.5. This simulation does not

catch the true change point which is  $k = 500$ . The reason is the same as other large sample size test results.

### 3.2.3.6 $r_1 = -0.2, r_2 = 0.4$

The difference between  $r_1$  and  $r_2$  is same as in the previous such in, but  $r_1$  and  $r_2$  have different signs.

When  $n_1 > n_2$ , the mean of the estimated change point is 43.56 with standard deviation 14.44. The median of the estimated change point is 46. This simulation caught the true change point which is  $k = 50$ .

For  $n_1 = n_2 = 25$ , the mean of the estimated change point is 27.86 with standard deviation 9.37. The median of the estimated change point is 28. This simulation caught the true change point which is  $k = 25$ . Under the condition of the number of observations setting  $n_1 = n_2$ , the results are identical in every  $r_1$  and  $r_2$  settings we examined.

For  $n_1 < n_2$ , the mean of the estimated change point is 27.28 with standard deviation 11.45. The median of the estimated change point is 28. This simulation does not catch the true change point which is  $k = 5$ . We did not get the meaningful results because of the same reason as other  $n_1 < n_2$  examples.

For large sample size with  $n_1 = 500, n_2 = 20$ , the mean of the estimated change point is 285.5 with standard deviation 116.62. The median of the estimated change point is 288. This simulation does not catch the true change point which is  $k = 500$ . The reason is the large difference between  $n_1$  and  $n_2$ , that does not satisfy the  $\frac{1}{2}n_1 \leq n_2 \leq n_1$ .

All in all, based on the analysis given above,  $\gamma_n$  does not seem to affect the results of the test. Under the  $r_1$  and  $r_2$  settings we examined, the method effectively caught the change point especially when the number of observations after the change point is equal to the number

of observations before the change point, which is  $n_1 = n_2$ . When  $\frac{1}{2}n_1 \leq n_2 < n_1$ , the proposed method also gave the meaningful results, but the accuracy and sensitivity of the test here is not as good as the previous  $n_1$  and  $n_2$  setting. That may due to the large difference between  $n_1$  and  $n_2$ , and results would be biased to the side with much more observations of data.

## CHAPTER 4

### APPLICATION TO REAL DATA

We selected Johnson & Johnson and Eli Lilly and Company two American global pharmaceutical companies' stock price, from April 1<sup>st</sup>, 2009 to April 15<sup>th</sup>, 2014 as our real dataset. They have similar macroeconomic factors. They share a similar industrial structure since both of them are pharmaceutical companies. These macroeconomic factors can affect the two companies' stock prices simultaneously. Hence, we suppose that they have a similar underlying economic environment. On the other hand, the companies' own situation and market factors also affect the stock prices. That's what we are looking for when a change point occurs via cross-correlation.

Johnson & Johnson is an American pharmaceutical manufacturer founded in 1886. Their product line focuses on pharmaceuticals, medical devices and diagnostics, as well as consumer products. Johnson & Johnson includes over 250 subsidiary companies with operations in over 57 countries. Their products, like Self-Measured Blood Glucose Monitors, Tylenol, Johnson's Baby and Clean & Clear, are sold in over 175 countries. In 2014, Johnson & Johnson had 74.331 billion US dollars in revenue. Johnson & Johnson's stock is traded on the New York Stock Exchange, code JNJ. Johnson & Johnson's stock is also part of the Dow Jones Industrial Average Index and S&P 500 Index. (The Wall Street Journal, 2015 and Johnson & Johnson, 2015)

Eli Lilly and Company is an American global pharmaceutical company established in 1876. Eli Lilly and Company was the first business to mass produce penicillin and insulin, and is also

the first pharmaceutical company to use recombinant DNA to produce human insulin. Eli Lilly and Company has offices in 17 other countries and more than 37 thousand employees as of 2013. Their products, Alimta for lung cancer, Prozac for antidepressants, Cialis for erectile dysfunction, and Humalog for Type 1 and Type 2 diabetes are sold in over 125 countries. As of 2013, Eli Lilly has 23.113 billion US dollars in revenue. Eli Lilly and Company's stock is traded on the New York Stock Exchange, code LLY. (The Wall Street Journal, 2015 and Eli Lilly and Company, 2015)

From Fig. 1.6, visually there might be two main structural change points around Jan. 1<sup>st</sup>, 2013 and April 1<sup>st</sup>, 2013. At first we used the full data to detect a main structural change point around Jan. 1<sup>st</sup>, 2013. We did the cross-correlation analysis first before and after Jan. 1<sup>st</sup>, 2013. Table 4.1 was created by the results of cross-correlation analyses. As we said before, most stock data is best modeled as AR (1) time series, which means we should set lag = -1, 0, 1 in our results.

Table 4.1 Cross-Correlation for the Full Dataset  
Due to the Visual Change Point, Jan. 1<sup>st</sup>, 2013

Before Jan. 1 <sup>st</sup> , 2013		After Jan. 1 <sup>st</sup> , 2013	
LAG	Cross-Correlation	LAG	Cross-Correlation
-1	0.770	-1	-0.001
0	0.776	0	0.015
1	0.768	1	-0.001

For the full dataset, before the visual change point, Jan. 1<sup>st</sup>, 2013, the cross-correlation is 0.776 with 0 time lag. That means LLY and JNJ's stock prices tend to move strongly together without lag or lead. After the visual change point, the cross-correlation is 0.015 with 0 time lag, which means the cross-correlation is so weak that we can say there is no relationship between these two stock prices. Hence there was a large difference between cross-correlation

before and after the visual change point. Since in the business calendar, Jan. 1<sup>st</sup>, 2013 was a Sunday without any stock trading, we can concluded that there may be a main structural change point on Jan. 2<sup>nd</sup>, 2013.

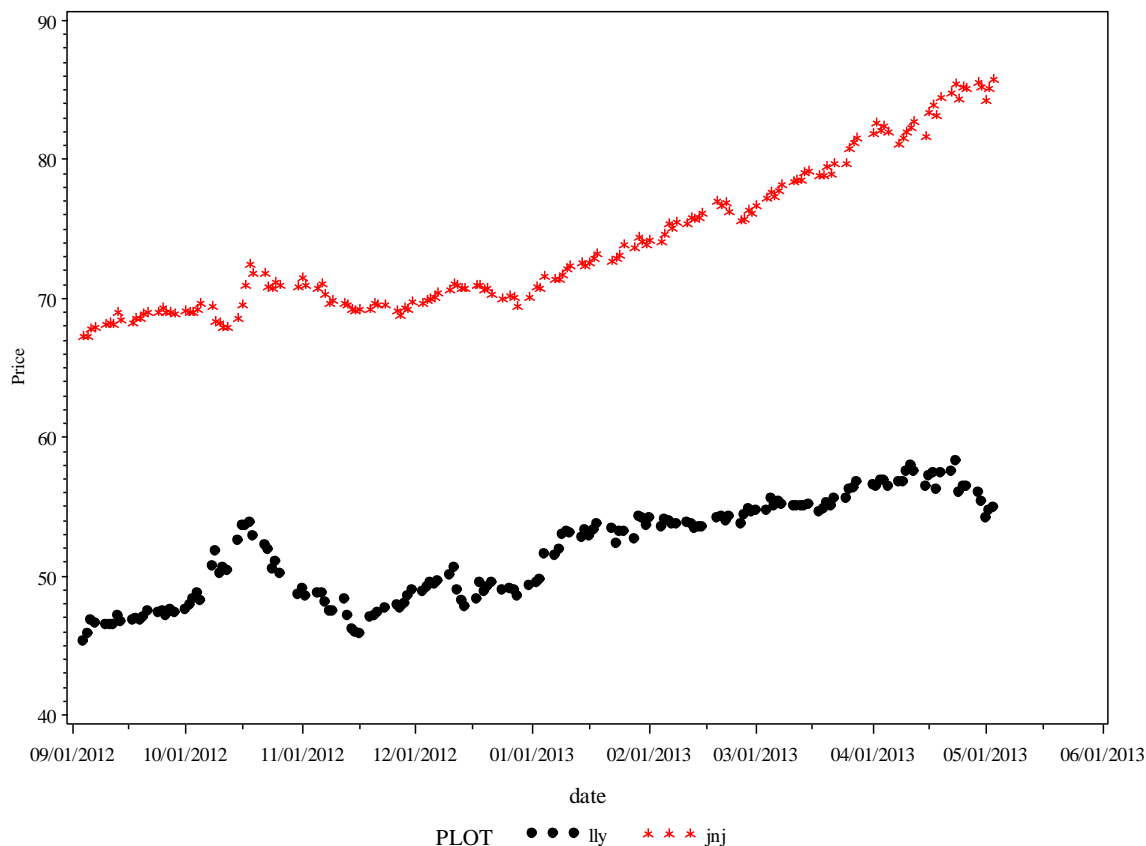
Then we used proposed method to test the full dataset. As we stated before,  $\gamma_n = 3$ . The observation number of estimated change point is 521, which is April 25<sup>th</sup>, 2011. The test did not catch either the visual main structural change point around Jan. 1<sup>st</sup>, 2013 or the one around April 1<sup>st</sup>, 2013. To analyze this result, we computed  $n_1$  and  $n_2$  in the full dataset; the number of observations before the visual change point is 946, which means  $n_1 = 946$ , the number of observations after the visual change point is 323, which means  $n_2 = 323$ . There was a large difference between  $n_1$  and  $n_2$ , which does not satisfy  $\frac{1}{2}n_1 \leq n_2 \leq n_1$  for which the proposed method works well, as determined by the simulation study. Hence the proposed method cannot catch the visual change point in the full dataset. In addition, we got this result with 323 observations after the visual main structural change point. This is almost 1.5 years after the visual main structural change point, and that does not satisfy our goal in this paper.

Table 4.2 Estimated Change Point for Different Time Intervals

Interval	Estimated Change Point	Change Point date
10.31.2012-1.18.2013, 9 weeks before, 2 weeks after	23	Nov. 30 <sup>th</sup> , 2012
10.31.2012-3.8.2013, 5 weeks after	42	Dec. 28 <sup>th</sup> , 2012
10.31.2012-5.3.2013, 17 weeks after	53	Jan. 15 <sup>th</sup> , 2013
9.4.2012-1.18.2013, 17 weeks before, 2 weeks after	24	Oct. 4 <sup>th</sup> , 2012
9.4.2012-3.17.2013, 11 weeks after	81	Dec. 28 <sup>th</sup> , 2012
9.4.2012-5.3.2013, 17 weeks after	83	Jan. 2 <sup>nd</sup> , 2013

In actual financial analysis, we do not typically use such a long a time interval, but a small moving window to analyze the data. Hence we tried some different reduced time windows to see how the results of real data match the simulation results. Table 4.2 shows part of the results of different time intervals in real data. The complete results are in Appendix C. In simulation, we concluded that our test effectively caught the change point especially well when the number of observations after the change point was equal to the number of observations before the change point, which is  $n_1 = n_2$ . So we selected data from September 4<sup>th</sup>, 2012 to May 3<sup>rd</sup>, 2013. In this case, the time window before and after the visual change point were both 3 business months, and a total of 163 observations. That is  $n_1 \approx n_2$ ,  $n = n_1 + n_2 = 163$ . Fig 4.1 is the plot of the stock price of LLY and JNJ in the selected date interval.

**Fig. 4.1** plot of LLY and JNJ stock prices from September 4th, 2012 to May 3rd, 2013



From the cross-correlation analysis we constructed the following Table 4.3 for the reduced dataset from September 4<sup>th</sup>, 2012 to May 3<sup>rd</sup>, 2013. The cross-correlation coefficient before the estimated change point is 0.5656 when lag = 0; it changed to 0.9126 when lag = 0. That means before the estimated main structural change point, the JNJ's stock price changes on the same day when the LLY's stock price changes, and the relationship between them is moderate. After the estimated change point, the JNJ's stock price changes on the same day as the LLY's stock price changes, and the relationship between them is very strong without lag or lead. We can conclude that there is an estimated main structural change point on Jan. 2<sup>nd</sup>, 2013 due to the cross-correlation change from 0.5656 with lag = 0 to 0.9126 with lag = 0.

Table 4.3 Cross-Correlation for the Reduced Dataset  
from September 4<sup>th</sup>, 2012 to May 3<sup>rd</sup>, 2013

Before Jan. 2 <sup>nd</sup> , 2013		After Jan. 2 <sup>nd</sup> , 2013	
LAG	Cross-Correlation	LAG	Cross-Correlation
-1	0.4378	-1	0.8355
0	0.5656	0	0.9126
1	0.5922	1	0.8978

From the test's result using the reduced data, the observation number of the estimated change point is 80, which is Jan. 2<sup>nd</sup>, 2013. The test caught the visual change point very well.

Then we reduced the time window to a new interval from Dec. 3<sup>rd</sup>, 2012 to Jan. 31<sup>st</sup>, 2013, which are one business month before and after the visual change point respectively. From the cross-correlation analysis we constructed the following Table 4.4 for the reduced dataset from Dec. 3<sup>rd</sup>, 2012 to Jan. 31<sup>st</sup>, 2013. The cross-correlation coefficient before the estimated change point is 0.149 when lag = 0; it changed to 0.773 when lag = 0. That means before the estimated main structural change point, the JNJ's stock price changes on the same day when the LLY's



stock price changes, and the relationship between them is weak. After the estimated main structural change point, the JNJ's stock price changes on the same day as the LLY's stock price changes, and the relationship between them is strong without lag or lead. We can also conclude that there is an estimated main structural change point on Jan. 3<sup>rd</sup>, 2013 due to the cross-correlation change from 0.149 with lag = 0 to 0.773 with lag = 0.

Table 4.4 Cross-Correlation for the Reduced Dataset  
from Dec. 3<sup>rd</sup>, 2012 to Jan. 31<sup>st</sup>, 2013

Before Jan. 3 <sup>rd</sup> , 2013		After Jan. 3 <sup>rd</sup> , 2013	
LAG	Cross-Correlation	LAG	Cross-Correlation
-1	0.070	-1	0.612
0	0.149	0	0.773
1	-0.176	1	0.548

From the result of test, the observation number of estimated change point is 22, which is Jan. 3<sup>rd</sup>, 2013. The test caught the visual change point very well.

We noticed that there also may be another main structural change point around April 1<sup>st</sup>, 2013 due to the trend of LLY and JNJ's stock prices. Fig 4.2 is the plot of the stock prices of LLY and JNJ from Jan. 2<sup>nd</sup>, 2013 to Jun. 28<sup>th</sup>, 2013, which are three months before and after the visual main structural change point, respectively. From the plot we found that there may be a main structural change point around beginning of the April, because before the point around April 1<sup>st</sup>, 2013, the stock prices of LLY and JNJ have trends of the same direction but different levels, after the point around April 1<sup>st</sup>, 2013, the stock prices of LLY and JNJ have trends with different directions and different levels.

We did cross-correlation analysis for selected time intervals due to the number of observation before the visual change point is almost equal to the number of observation after

the visual change point, and created Table 4.5. In Table 4.5, the  $r_1$  and  $r_2$  of selected time intervals are very different. We can conclude that there is an estimated main structural change point around April 1<sup>st</sup>, 2013 due to the cross-correlation changes.

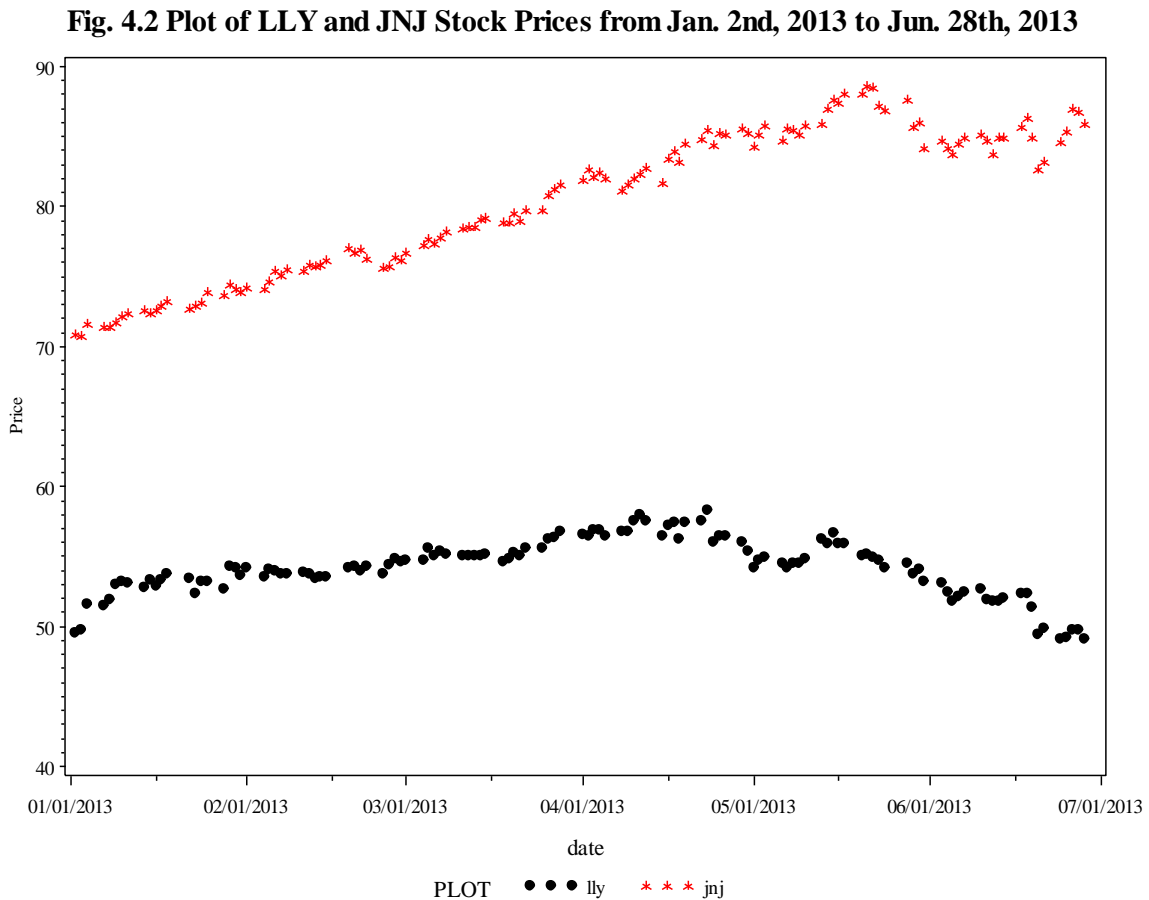


Table 4.5 Cross-Correlation Analysis for Different Time Intervals around April 1<sup>st</sup>, 2013

Time Interval	Time Lag	$r_1$	$r_2$
1.2.2013-6.28.2013	-1	0.833	-0.195
	0	0.899	-0.187
	1	0.780	-0.228
2.1.2013-5.31.2013	-1	0.798	-0.476
	0	0.906	-0.485
	1	0.817	-0.630
3.1.2013-4.30.2013	-1	0.638	-0.126
	0	0.842	-0.130
	1	0.639	-0.391

We did the test to detect this visual change point by selected different time intervals. Table 4.6 shows the results of test in different time intervals. From the Table. 4.6, the proposed method found that the location of change points depends on data size. We could not conclude that there is one estimated main structural change point around April 1<sup>st</sup>, 2013. The different estimated change points in selected time interval from Jan. 4<sup>th</sup>, 2013 to Jun. 25<sup>th</sup>, 2013 may due to multiple change points occur in the selected time interval, or may due to the biased number of observations of data before and after the true change point in the selected time intervals.

Table 4.6 Estimated Change Point for Different Time Intervals around April 1<sup>st</sup>, 2013

Time Interval	Estimated Change Point	Estimated Change Point date
1.2.2013-6.28.2013	42	3.4.2013
2.1.2013-5.31.2013	32	3.19.2013
3.1.2013-4.30.2013	17	3.25.2013

## CHAPTER 5

### SUMMARY AND DISCUSSION

For time series data, change point theory is a mathematical methodology that focuses on detecting a particular time point when data flow changes. Industry uses it on quality control. Economists use this theory to study the stock market. Many methodologies have been built on this theory. A maximum likelihood estimator was introduced by P.K. Bhattacharya (P.K. Bhattacharya, 1987). Bayesian methods have been widely studied (Tamura, 2008 and Kleykamp, 2004).

In this study, we wanted to know how one stock price changes relative to another one and how quickly we can detect a change point via cross-correlation. We used a test to find the change point in two-group time series data. We analyzed the cross-correlation on subgroups, before and after the estimated change point. We focus on how the cross-correlation changes before and after the change point between the two groups of data to detect the change point.

A simulation study is presented in Chapter 3. The parameters are set at different situations. We simulated the change point shown at different parts over time: at the early stage, at the middle stage and at the later stage and with different sets of cross-correlation. Our simulation results have shown that the proposed method works effectively when the change point is shown at the middle stage and the later stage of time with cross-correlation setting we examined.

We examined the cross-correlations between JNJ and LLY. Then we used a statistical test proposed by Dominik Wied (2011) to detect the change point. This proposed method is applied

to real stock data of JNJ and LLY for the full dataset from April, 2009 to April, 2014, the reduced dataset from September, 2012 to the beginning of May, 2013 and different reduced time intervals from Jan. 2<sup>nd</sup>, 2013 to Jun. 28<sup>th</sup>, 2013.

For this test proposed by Dominik Wied (2011), we found that the bandwidth  $\gamma_n$  did not seem to affect the results of test. Any reasonable bandwidth  $\gamma_n$  could be chosen in the test. We set  $\gamma_n = 3$  in our study. The number of observations before and after the true change point affect the results of the test. Meaningful results can be obtained when the number of observations after the change point is equal to or greater than half of the number of observations before the change point and less than or equal to the number of observations before the change point, which is  $\frac{1}{2}n_1 \leq n_2 \leq n_1$ . With this criterion, the method can effectively catch the change point especially when  $n_1 = n_2$ . On the other hand, the cross-correlation settings we looked at did not affect the results of test.

The strengths of the proposed method are that it can effectively determine the change point; the criteria of this testing is clear; the result is easy to interpret; and this methodology is easy to apply to time series data with two or more groups. It is suggested that any reasonable  $\gamma_n$  can be used in the test. Also using a shorter moving time window to detect the change point may also be more useful than a whole historical dataset. The challenge of the proposed method is that it is not effective at catching the change point when  $|n_1 - n_2| > \Delta$  with large  $\Delta$ . It may be due to a lack of observations in subsample  $n_1$ , which represents the observations before the change point, or the lack of observations in subsample  $n_2$ , which represents the observations after the change point. In addition, when there are more than two change points in the dataset, the proposed method may not be effective at catching the main structural change point. Future research can be expanded to more than two-groups, higher dimension data. It is also of interest

to derive the power of the proposed method. Dose the proposed method have the same power in every  $r_1$  and  $r_2$  setting? Finally, detecting more than two main structural change points can also be a direction of future study.

#### REFERENCES

- Everitt, B.S. and A. Skrondal, *THE CAMBRIDGE DICTIONARY OF STATISTICS*, ed. 4. 2010: Cambridge University Press.
- Hofrichter, J., *CHANGE POINT DETECTION IN GENERALIZED LINEAR MODELS*. Dissertation paper, 2007.
- Pettitt, A.N., *A NON-PARAMETRIC APPROACH TO THE CHANGE-POINT PROBLEM*. Journal of the Royal Statistical Society. Series C (Applied Statistics), 1979. **28**(2): p. 126-135.
- Barry, D. and J.A. Hartigan, *A BAYESIAN ANALYSIS FOR CHANGE POINT PROBLEMS*. Journal of the American Statistical Association, 1993. **88**(421): p. 309-319.
- Tamura, N., e. Yuge, and S. Yanagi, *ESTIMATION OF THE CHANGE POINT FOR FAILURE-CENSORED DATA VIA BAYESIAN INFORMATION CRITERION*, in *Secure System Integration and Reliability Improvement, 2008. SSIRI '08. Second International Conference*. 2008. p. 207-208.
- Sur, M., et al., *CHANGE DETECTION IN ANIMAL MOVEMENT USING DISCRETE WAVELET ANALYSIS*. ECOLOGICAL INFORMATICS, 2014. **20**: p. 47-57.
- Wied, D., et al., *A FLUCTUATION TEST FOR CONSTANT SPEARMAN'S RHO*. SFB 823 Discussion paper, 2011. **2011/16**: p. 2011a.
- Papoulis, A., *THE FOURIER INTEGRAL AND ITS APPLICATIONS*. 1962: New York: McGraw-Hill. 244-245 and 252-253.

Bracewell, R., *PENTAGRAM NOTATION FOR CROSS CORRELATION. THE FOURIER TRANSFORM AND ITS APPLICATIONS.* . 1965: New York: McGraw-Hill.

Rodgers, J.L. and W.A. Nicewander, *THIRTEEN WAYS TO LOOK AT THE CORRELATION COEFFICIENT.* The American Statistician, 1988. **42**(1): p. 59-66.

Tsay, R.s., *ANALYSIS OF FINANCIAL TIME SERIES*, ed. r. edition. 2010: John wiley & sons, INC., publication.

Aarts, R.M., R. Irwan, and A.J.E.M. Janssen, *EFFICIENT TRACKING OF THE CROSS-CORRELATION COEFFICIENT.* Ieee Transactions On Speech And Audio Processing, 2002. **10**(6): p. 391-402.

Liu, L., *CROSS-CORRELATIONS BETWEEN CRUDE OIL AND AGRICULTURAL COMMODITY MARKETS.* Physica A, 2014. **395**: p. 293-302.

Alaoui, M.E. and S. Benbachir, *MULTIFRACTAL DETRENDED CROSS-CORRELATION ANALYSIS IN THE MENA AREA.* Physica A, 2013. **392**: p. 5985-5993.

Podobnik, B., et al., *CROSS-CORRELATIONS BETWEEN VOLUME CHANGE AND PRICE CHANGE.* PNAS, 2009. **106**(52): p. 22079–22084.

Andrews, D.W.K., *HETEROSKEDASTICITY AND AUTOCORRELATION CONSISTENT COVARIANCE MATRIX ESTIMATION.* Econometrica, 1991. **59**(3): p. 817-858.

Journal, T.W.S. *JOHNSON & JOHNSON.* 2015; Available from: <http://quotes.wsj.com/JNJ>.

Johnson, J. *JOHNSON & JOHNSON.* 2015; Available from: <http://www.jnj.com/>.

Journal, T.W.S. *ELI LILLY AND COMPANY.* 2015; Available from: <http://quotes.wsj.com/LLY>.

Company, E.L.a. *ELI LILLY AND COMPANY* 2015; Available from: <http://www.lilly.com/Pages/Home.aspx>.

Bhattacharya, P.K., *MAXIMUM LIKELIHOOD ESTIMATION OF A CHANGE-POINT IN THE DISTRIBUTION OF INDEPENDENT RANDOM VARIABLES: GENERAL MULTIPARA METER CASE*. Journal of Multivariate Analysis, 1987. **23**(2): p. 183-208.

Western, B. and M. Kleykamp, *A BAYESIAN CHANGE POINT MODEL FOR HISTORICAL TIME SERIES ANALYSIS*. Political Analysis, 2004. **12**(4): p. 354-374.



## APPENDIX A

### Code for Simulations

FORTRAN code:

```

program simul
implicit none
real ::
r1,r2,c1,c2,e,pa,pb,sumpc,sumpm,gamma_n,ker,hd,d,dhatp,dhat,pbsq,kf,sumpuj,nnn,maxpd,
w,Maxij
real :: yy1,yy2,e1
integer, parameter :: NN = 1000
integer, parameter :: pi = 3.1415926
real,dimension(NN) :: xp,y,xr,yr,pu,rhok,pd,puj,cp
integer :: i,n1,n2,n,j,m,k,nrep,ij,gn
integer,dimension(1) :: place
character(len=16) :: filename
print *,'Input numbers before change point to continue'
read *, n1
print *,'Input numbers after change point to continue'
read *, n2
print *,'Input coefficient of relationship before change point to continue'
read *, r1
print *,'Input coefficient of relationship after change point to continue'
read *, r2
print *,'Gamma_n can be<=',int(sqrt(float(n1+n2)))
print *,'Input gamma_n'
read *, gamma_n
print *,'Input the output file name'
read *,filename
open(unit=3, file='C:\Users\Administrator\Desktop\results\//filename//.txt')

```

```
c1 = sqrt(1-r1**2)
c2 = sqrt(1-r2**2)
call random_seed()
d=2
do nrep=1,100
    Maxij=int(sqrt(float(n1+n2)))
    Do ij=2,Maxij
        Gamma_n = ij
yy1=0
yy2=0
pa=0
pb=0
sumpc=0
sumpm=0
ker=0
hd=0
d=0
dhatp=0
dhat=0
pbsq=0
kf=0
sumpuj=0
nnn=0
maxpd=0
w=0
xp=0
y=0
xr=0
yr=0
pu=0
rhok=0
```

```

pd=0
puj=0
cp=0
place=0
y=0
  do i = 1,n1+n2
call random_number(e)
call random_number(e1)
yy1 = sqrt(-2.0*log(e))*cos(2*e1*pi)
yy2 = sqrt(-2.0*log(e))*sin(2*e1*pi)
xp(i)=yy1
  if (i<n1) then
    y(i)=xp(i)*r1+yy2*c1
  else
    y(i)=xp(i)*r2+yy2*c2
  endif
end do
n=n1+n2
call rank(xp,n,xr)
call rank(y,n,yr)
pa=0
do i = 1,n1+n2
pu(i)=(1-xr(i)/n)*(1-yr(i)/n)
pa=pa+(((1-xr(i)/n)**2)*((1-yr(i)/n)**2))/n
pb=(sum(pu))/n
  end do
  pbsq=pb**2
  sumpm=0
do m=1,n-1
  ker = m/gamma_n
  if (ker<=1) then

```

```

    kf=1
else
    kf=0
end if
sumpc=0
do j=1,n-m
    sumpc=sumpc+((1-xr(j)/n)*(1-yr(j)/n)*(1-xr(j+m)/n)*(1-yr(j+m)/n))/n
end do
do gn=1,gamma_n
    sumpm = sumpm + kf*(sumpc-pbsq)
end do
end do
hd = 3 ! because hd=(d+1)/(2**d-(d+1)) where d=2
dhatp=(hd**2)*(2**(2*d))*(pa-pbsq+2*sumpm)
if(dhatp.le.0) dhatp=1
dhat=1/(sqrt(dhatp))
do k=1,n
    puj(j)=0
    do j=1,k
        puj(j)=(1-xr(j)/n)*(1-yr(j)/n)
    end do
    sumpuj=(sum(puj)*4/k-1)*3
    nnn=n
    rhok(k)=sumpuj
end do
do k=1,n
    pd(k)=(k/sqrt(nnn))*(rhok(k)-rhok(n))
end do
maxpd=maxval(abs(pd(1:n1+n2)))
w=dhat*maxpd
place=maxloc(abs(pd(1:n1+n2)))

```

```

cp(nrep)=place(1)
write(3,*) n1,n2,r1,r2,gamma_N,w,cp(nrep)
end do
end do
close (3)
end program simul

!+++++
SUBROUTINE RANK(X,N,XR)
! DOWNLANDED FROM HTTP://WWW.NIST.GOV/
  DIMENSION X(N),XR(N)
  COMMON /BLOCK4/ XS(1000)
  AN=N
  IPR=6
  HOLD=X(1)
  DO 60 I=2,N
    IF(X(I).NE.HOLD)GOTO 90
60 CONTINUE
  WRITE(IPR, *)HOLD
  AVRANK=(AN+1.0)/2.0
  DO I=1,N
    XR(I)=AVRANK
  end do
90 CONTINUE
  CALL SORT(X,N,XS)
  NM1=N-1
  XPREV=X(1)
  DO 700 I=1,N
    JMIN=1
    IF(X(I).GT.XPREV)GOTO 770
    IF(I.EQ.1)GOTO 790
    IF(X(I).EQ.XPREV)GOTO 750

```

```
GOTO 790
750 CONTINUE
  XR(I)=RPREV
  GOTO 880
770 CONTINUE
  JMIN=K
  IF(JMIN.LT.N)GOTO 790
  IF(JMIN.EQ.N)GOTO 820
  IBRAN=1
  WRITE(IPR,*)IBRAN
  WRITE(IPR,*)JMIN
  STOP
790 CONTINUE
  DO 800 J=JMIN,NM1
  IF(X(I).NE.XS(J))GOTO 800
  JP1=J+1
  DO 900 K=JP1,N
  IF(XS(K).NE.XS(J))GOTO 950
900 CONTINUE
  K=N+1
950 CONTINUE
  AVRANK=J+K-1
  AVRANK=AVRANK/2.0
  XR(I)=AVRANK
  GOTO 880
800 CONTINUE
820 CONTINUE
  J=N
  K=N+1
  IF(X(I).EQ.XS(J))GOTO 850
  IBRAN=2
```

```

WRITE(IPR,*)IBRAN
WRITE(IPR,*)X(I),XS(J)
STOP
850 CONTINUE
XR(I)=N
880 CONTINUE
XPREV=X(I)
RPREV=XR(I)
700 CONTINUE
RETURN
END subroutine rank
!+++++
SUBROUTINE SORT(X,N,Y)
! DOWNLANDED FROM HTTP://WWW.NIST.GOV/
DIMENSION X(N),Y(N)
integer,DIMENSION(1000):: IU,IL
IPR=6
HOLD=X(1)
DO 60 I=2,N
IF(X(I).NE.HOLD)GOTO 90
60 CONTINUE
WRITE(IPR,*)HOLD
DO 61 I=1,N
Y(I)=X(I)
61 CONTINUE
RETURN
90 CONTINUE
DO 100 I=1,N
Y(I)=X(I)
100 CONTINUE
NM1=N-1

```

```
DO 200 I=1,NM1
  IP1=I+1
  IF(Y(I).LE.Y(IP1))GOTO 200
  GOTO 250
200 CONTINUE
  RETURN
250 M=1
  I=1
  J=N
305 IF(I.GE.J)GOTO 370
310 K=I
  MID=(I+J)/2
  AMED=Y(MID)
  IF(Y(I).LE.AMED)GOTO 320
  Y(MID)=Y(I)
  Y(I)=AMED
  AMED=Y(MID)
320 L=J
  IF(Y(J).GE.AMED)GOTO 340
  Y(MID)=Y(J)
  Y(J)=AMED
  AMED=Y(MID)
  IF(Y(I).LE.AMED)GOTO 340
  Y(MID)=Y(I)
  Y(I)=AMED
  AMED=Y(MID)
  GOTO 340
330 Y(L)=Y(K)
  Y(K)=TT
340 L=L-1
  IF(Y(L).GT.AMED)GOTO 340
```



```
      TT=Y(L)
350 K=K+1
      IF(Y(K).LT.AMED)GOTO 350
      IF(K.LE.L)GOTO 330
      LMI=L-I
      JMK=J-K
      IF(LMI.LE.JMK)GOTO 360
      IL(M)=I
      IU(M)=L
      I=K
      M=M+1
      GOTO 380
360 IL(M)=K
      IU(M)=J
      J=L
      M=M+1
      GOTO 380
370 M=M-1
      IF(M.EQ.0)RETURN
      I=IL(M)
      J=IU(M)
380 JMI=J-I
      IF(JMI.GE.11)GOTO 310
      IF(I.EQ.1)GOTO 305
      I=I-1
390 I=I+1
      IF(I.EQ.J)GOTO 370
      AMED=Y(I+1)
      IF(Y(I).LE.AMED)GOTO 390
      K=I
395 Y(K+1)=Y(K)
```

```

K=K-1
IF(AMED.LT.Y(K))GOTO 395
Y(K+1)=AMED
GOTO 390
END subroutine sort

```

SAS code:

```

%let dirname = C:\Users\Administrator\Desktop\results;
filename DIRLIST pipe "dir /B &dirname\*.txt";
data dirlist ;
    length fname $256;
    infile dirlist length=reclen ;
    input fname $varying256. reclen ;
run;
data results (drop=fname);
    length myfilename $100;
    set dirlist;
    filepath = "&dirname\"||fname;
    infile dummy filevar = filepath length=reclen end=done missover;
    do while(not done);
        myfilename = filepath;
        input n1 n2 r1 r2 gamma_n w cp;
        output;
    end;
run; /*http://www.ats.ucla.edu/stat/sas/faq/multi\_file\_read.htm*/

ods rtf;
data r1eq1;
set results;
where r1=1 and gamma_n=3;
run;
proc univariate data=r1eq1 normaltest loccount plots;
class n1 n2;
var cp;
run;
proc means data=r1eq1 maxdec=2 mean std p25 p50 p75 ;
class n1 n2 gamma_n;
var cp;
output
out = r11 gamma_n;
title "r1=1 r2=-1";
run;

```

```

data r1eq05;
set results;
where r1=0.5 and gamma_n=3;
run;
proc univariate data=r1eq05 loccount plots normaltest;
class n1 n2;
var cp;
run;
proc means data=r1eq05 maxdec=2 mean std p25 p50 p75 ;
class n1 n2 gamma_n;
var cp;
output
out = r105gamma_n;
title "r1=0.5 r2=-0.5";
run;
data r1eqn07;
set results;
where r1=-0.7 and gamma_n=3;
run;
proc univariate data=r1eqn07 loccount plots normaltest;
class n1 n2 gamma_n;
var cp ;
run;
proc means data=r1eqn07 maxdec=2 mean std p25 p50 p75 ;
class n1 n2 gamma_n;
var cp;
output
out = r1n07gamma_n;
title "r1=-0.7 r2=0.4";
run;
data r1eq04;
set results;
where r1=0.4 and gamma_n=3;
run;
proc univariate data=r1eq04 normaltest loccount plots;
class n1 n2;
var cp;
run;
proc means data=r1eq04 maxdec=2 mean std p25 p50 p75 ;
class n1 n2 gamma_n;
var cp;
output
out = r104gamma_n;
title "r1=0.4 r2=0.7";
run;
data r1eq02;

```

```

set results;
where r1=0.2 and gamma_n=3;
run;
proc univariate data=r1eq02 normaltest loccount plots;
class n1 n2;
var cp;
run;
proc means data=r1eq02 maxdec=2 mean std p25 p50 p75 ;
class n1 n2 gamma_n;
var cp;
output
out = r102gamma_n;
title "r1=0.2 r2=0.8";
run;
data r1eqn04;
set results;
where r1=-0.4 and gamma_n=3;
run;
proc univariate data=r1eqn04 normaltest loccount plots;
class n1 n2;
var cp;
run;
proc means data=r1eqn04 maxdec=2 mean std p25 p50 p75 ;
class n1 n2 gamma_n;
var cp ;
output
out = r1n04gamma_n;
title "r1=-0.4 r2=0.2";
run;
ods rtf close;

```

APPENDIX B  
Results of Simulations

For  $r_1 = 1, r_2 = -1$

Analysis Variable : cp								
n1	n2	gamma_n	N Obs	Mean	Std Dev	25th Pctl	50th Pctl	75th Pctl
5	5	3	1000	3.25	1.33	2.00	3.00	4.00
	25	3	1000	7.00	5.75	3.00	4.00	9.00
	50	3	1000	12.89	11.82	4.00	7.50	20.00
25	5	3	1000	16.23	5.89	12.00	17.00	22.00
	25	3	1000	21.24	3.91	20.00	23.00	24.00
	50	3	1000	22.18	3.61	21.00	23.00	24.00
50	5	3	1000	30.23	12.07	21.00	31.00	41.00
	25	3	1000	42.88	7.35	40.00	46.00	48.00
	50	3	1000	45.73	4.81	45.00	47.00	49.00
200	5	3	1000	104.52	44.89	68.00	105.00	138.00
	10	3	1000	121.39	48.47	84.00	125.00	164.00
	20	3	1000	153.22	44.71	124.50	170.00	191.00
	50	3	1000	178.69	25.42	170.00	189.00	197.00
500	5	3	1000	254.40	110.09	171.00	253.00	341.50
	10	3	1000	272.11	116.55	177.00	272.00	367.00
	20	3	1000	329.38	123.43	231.50	350.00	440.00
	50	3	1000	433.38	77.40	408.00	464.00	490.00

For  $r_1 = 0.5, r_2 = -0.5$

Analysis Variable : cp								
n1	n2	gamma_n	N Obs	Mean	Std Dev	25th Pctl	50th Pctl	75th Pctl
5	5	3	1000	4.25	2.07	3.00	4.00	6.00
	25	3	1000	12.40	7.01	6.00	12.00	18.00
	50	3	1000	23.97	12.89	14.00	24.00	34.00
25	5	3	1000	14.62	6.15	10.00	15.00	20.00
	25	3	1000	21.77	7.95	17.00	22.00	26.00
	50	3	1000	28.21	12.48	20.00	25.00	35.00
50	5	3	1000	27.39	11.56	18.00	28.00	37.00
	25	3	1000	37.95	12.31	30.00	41.00	48.00
	50	3	1000	44.52	12.98	37.00	46.00	50.00
200	5	3	1000	101.81	44.76	65.00	100.00	139.00
	10	3	1000	107.72	46.77	69.00	109.00	145.50
	20	3	1000	124.60	49.70	86.00	129.00	166.00
	50	3	1000	158.81	44.48	133.00	175.00	195.00
500	5	3	1000	252.68	110.43	164.00	252.00	341.00
	10	3	1000	248.87	113.32	153.00	249.00	341.00
	20	3	1000	275.70	117.52	184.50	278.50	372.00
	50	3	1000	357.82	119.61	268.00	387.00	464.00

For  $r_1 = -0.7, r_2 = 0.4$

Analysis Variable : cp								
n1	n2	gamma_n	N Obs	Mean	Std Dev	25th Pctl	50th Pctl	75th Pctl
5	5	3	1000	5.72	1.90	4.00	6.00	7.00
	25	3	1000	15.31	6.14	10.00	16.00	20.00
	50	3	1000	28.27	11.85	18.00	29.00	38.00
25	5	3	1000	19.14	6.69	14.00	21.00	24.00
	25	3	1000	28.09	7.35	24.00	27.00	32.00

Analysis Variable : cp								
n1	n2	gamma_n	N Obs	Mean	Std Dev	25th Pctl	50th Pctl	75th Pctl
	50	3	1000	35.69	12.38	26.00	32.00	44.00
50	5	3	1000	33.52	13.45	23.00	34.50	47.00
	25	3	1000	47.71	10.86	44.00	49.00	53.00
	50	3	1000	54.79	11.59	49.00	52.00	60.00
200	5	3	1000	116.19	50.47	75.50	116.00	158.00
	10	3	1000	132.96	51.30	94.00	137.00	180.00
	20	3	1000	156.88	47.57	129.00	173.00	197.00
	50	3	1000	187.16	27.46	179.50	197.00	202.00
500	5	3	1000	269.74	118.06	172.00	267.50	365.00
	10	3	1000	298.76	127.90	193.00	306.00	412.00
	20	3	1000	349.54	125.64	253.00	374.50	463.00
	50	3	1000	437.64	87.51	412.50	474.00	498.00

For  $r_1 = 0.4, r_2 = 0.7$

Analysis Variable : cp								
n1	n2	gamma_n	N Obs	Mean	Std Dev	25th Pctl	50th Pctl	75th Pctl
5	5	3	1000	5.05	2.13	3.00	5.00	7.00
	25	3	1000	15.25	6.53	10.00	15.00	20.00
	50	3	1000	27.96	11.76	19.00	28.00	38.00
25	5	3	1000	15.76	6.71	10.00	16.00	21.00
	25	3	1000	26.02	10.54	18.00	26.00	35.00
	50	3	1000	37.56	16.17	24.00	37.00	50.00
50	5	3	1000	28.66	11.67	20.00	28.00	38.00
	25	3	1000	38.86	16.51	25.50	40.00	52.00
	50	3	1000	52.39	20.08	38.00	52.00	67.00
	5	3	1000	101.33	44.63	67.00	99.00	138.00

Analysis Variable : cp								
n1	n2	gamma_n	N Obs	Mean	Std Dev	25th Pctl	50th Pctl	75th Pctl
200	10	3	1000	108.45	47.09	71.00	112.00	145.00
	20	3	1000	113.72	50.36	74.50	115.00	153.00
	50	3	1000	132.49	55.08	90.00	132.00	181.00
500	5	3	1000	250.90	110.35	161.50	253.00	344.00
	10	3	1000	254.19	113.84	164.00	255.00	347.00
	20	3	1000	258.13	115.79	161.00	257.00	350.50
	50	3	2000	294.63	124.61	192.50	302.00	397.00

For  $r_1 = 0.2, r_2 = 0.8$

Analysis Variable : cp								
n1	n2	gamma_n	N Obs	Mean	Std Dev	25th Pctl	50th Pctl	75th Pctl
5	5	3	1000	5.25	2.04	4.00	5.00	7.00
	25	3	1000	15.43	6.00	11.00	15.00	20.00
	50	3	1000	28.21	11.60	19.00	28.00	37.00
25	5	3	1000	15.61	6.60	11.00	16.00	21.00
	25	3	1000	27.11	10.00	20.00	27.00	34.00
	50	3	1000	37.89	15.22	26.00	38.00	50.00
50	5	3	1000	28.95	12.26	20.00	29.00	39.00
	25	3	1000	40.97	15.97	29.00	43.00	53.00
	50	3	1000	53.90	18.90	42.00	53.00	68.00
200	5	3	1000	105.58	46.53	68.00	105.00	144.00
	10	3	1000	110.62	47.51	71.00	112.00	147.00
	20	3	1000	120.68	50.46	78.00	121.00	162.50
	50	3	1000	148.44	54.40	108.00	159.00	196.00
500	5	3	1000	248.06	112.92	153.00	248.00	337.00
	10	3	1000	262.66	116.49	167.50	261.50	354.00



Analysis Variable : cp								
n1	n2	gamma_n	N Obs	Mean	Std Dev	25th Pctl	50th Pctl	75th Pctl
	20	3	1000	278.50	119.69	179.00	278.00	374.00
	50	3	1000	331.45	128.64	228.00	345.00	448.00

For  $r_1 = -0.2, r_2 = 0.4$

Analysis Variable : cp								
n1	n2	gamma_n	N Obs	Mean	Std Dev	25th Pctl	50th Pctl	75th Pctl
5	5	3	1000	5.43	2.07	4.00	6.00	7.00
	25	3	1000	15.69	6.19	11.00	16.00	20.00
	50	3	1000	27.78	11.45	19.00	28.00	37.00
25	5	3	1000	16.79	6.66	12.00	17.00	22.00
	25	3	1000	27.86	9.37	22.00	28.00	35.00
	50	3	1000	38.43	14.72	26.50	38.00	50.00
50	5	3	1000	29.92	12.77	20.00	30.00	41.00
	25	3	1000	43.56	14.44	33.00	46.00	54.00
	50	3	1000	54.14	17.47	44.00	54.00	67.00
200	5	3	1000	107.62	47.63	70.00	105.00	147.00
	10	3	1000	114.11	49.46	73.50	114.00	156.00
	20	3	1000	127.28	52.63	86.00	132.00	172.00
	50	3	1000	157.06	51.33	122.00	172.00	199.00
500	5	3	1000	256.09	111.24	170.00	251.50	344.00
	10	3	1000	262.46	117.40	166.00	266.00	357.00
	20	3	1000	285.50	116.62	197.00	288.00	379.00
	50	3	1000	349.25	128.26	253.50	376.00	466.00

## APPENDIX C

### Results of Different Time Intervals

Date interval	Estimated date of change
10.31.2012-1.18.2013, 9 weeks before, 2 weeks after	11.30.2012
10.31.2012-2.1.2013, 4 weeks after	12.3.2012
10.31.2012-2.8.2013, 5 weeks after	12.28.2012
10.31.2012-2.15.2013, 6 weeks after	12.28.2012
10.31.2012-3.1.2013, 8 weeks after	12.28.2012
10.31.2012-3.28.2013, 12 weeks after	1.2.2013
10.31.2012-5.3.2013, 17 weeks after	1.15.2013
9.4.2012-1.18.2013, 17 weeks before, 2 weeks after	10.4.2012
9.4.2012-2.1.2013, 4 weeks after	10.4.2012
9.4.2012-3.1.2013, 8 weeks after	12.3.2012
9.4.2012-3.15.2013, 10 weeks after	12.5.2012
9.4.2012-3.22.2013, 11 weeks after	12.28.2012
9.4.2012-3.28.2013, 12 weeks after	12.28.2012

9.4.2012-5.3.2013, 17 weeks after	1.2.2013
--------------------------------------	----------