The Development of a Method for the Composite Coincident Indicator (CCI) for the United States

Chandra Putcha[1], Brian W. Sloboda[2],
[1]California State University at Fullerton, 800 N State College Blvd, Fullerton, CA 92831
[2]University of Phoenix, College of Doctoral Studies, 4025 S. Riverpoint Pkwy, Phoenix, AZ 85040

Abstract
The purpose of this paper is to continue work on a composite coincident indicator (CCI) for the United States and by each state as covered in Putcha and Sloboda (2017, 2018). This research looks at the economic time series for the United States by examining behavior of the key time series and applying the peak-valley algorithm as proposed by Schneider (2011) before creating the CCI. The detection of peaks and valleys in time series has been a longstanding problem in economic time series. To identify the trends in the economic time series, we provide two approaches to determine the trend in the time series: the geometric approach and the statistical definition of peaks and valleys. These two approaches can detect the significant trends within economic time series. This preliminary research examines the proposed variables that could be included into an eventual CCI for the United States.

Key Words: Peak-Valley Algorithm, local maximum and minimum, global maximum and minimum, time series

1. Introduction

The composite coincidental index (CCI) is an index combining two or more economic and financial variables. However, the process of combining two or more variables to develop a CCI can be tedious because of the behaviors underlying each of the variables being combined. As a remedy, peaks and valleys denote significant events in an economic time series. In fact, these events can be described as an abrupt increase with a recession or other economic event. A peak or a valley represents a significant event within a time series, and a significant event is a point where the function changes from increasing (decreasing) behavior to decreasing (increasing) behavior. Because of the latter changes, the identification of these behaviors is important for analysis, especially in the creation of the CCI. However, the detection of the peaks and valleys is not simple, and Schneider (2011) proposed an algorithm to determine the peaks and valleys in the time series data. In this preliminary research, we applied the Schneider (2011) algorithm to assist in the construction of a potential CCI. The literature concerning the historical development and recent advances in the development of the CCI is not presented here, but the reader is advised to review Putcha and Sloboda (2017, 2018) as well as other literature.

2. Methodology

2.1 Background on the Peaks and Valleys

To understand the behavior of the economic time series data, the time series were then analyzed for detecting local peaks after a change in the time series (e.g., a recession or growth). We employed a peak-valley detection algorithm as proposed by Schneider (2011) to detect the 'adaptation stages' by finding peak values that were above the global mean (average of the mean of all-time series amplitudes), which were separated from those stages that the time series spent minimal time on (peak values below global mean). This framework for peak-valley detection can also be of high value where the observational data are large or the observation time span is longer, and thus, it is difficult to observe the local maxima or the local minima in the data. The discussion sketches the background in the application of this algorithm proposed by Schneider (2011).

Mathematically, a peak of a given series means local maxima has a change in the slope from positive to negative.' Let $f(x)$ is a function which transforms x from a user-defined subdomain $A \subseteq R$ to the domain R as follows

$$f: A \rightarrow R$$

Let's consider an interval I=*(a, b),* and let's assume $I \cap A = \emptyset$

A local maximum is detected at point $x_0 \in I$ if $f(x_0) \geq f(x), \forall x \in I$. As for the local minimum, it is detected at point point $x_0 \in I$ if $f(x_0) \leq f(x), \forall x \in I$.

The difference between a global maximum and a local maximum is the domain of I. If $I \cap A = A$, then we will have a global maximum. It will also be similar for a global minimum and local minimum. Using the previous definitions for the global maximum and minimum, we obtain a global maximum point at x_0 if $f(x_0) \geq f(x)\ \forall x \in A$. Similarly, the function has a global minimum point at x_0 if $f(x_0) \leq f(x)\ \forall x \in A$. Based on the previous definitions, a peak is considered a local maximum, and a valley is a local minimum.

Given the latter definitions, the algorithms of detection of peaks and valleys must fulfill some additional requirements. That is, we need to assume that the time series is represented by a real function because this requirement guarantees there exist points between any two given points of a function. The latter requirement is called the continuity principle. From calculus, one can recall the concept of derivative using tangent lines. A tangent defines a linear slope which contacts a given function f(x) in each point x_0. Then, the angle between the tangent and the horizontal axis is used to describe the behavior of the function f(x) in point x_0. The value of the angle defines the slope of the tangent at $f(x_0)$. A positive value denotes an increasing trend of the function f(x) in point x_0, while a negative angle denotes a decreasing trend. If the angle equals zero, there is a flat trend in point x_0 which means there is a local extremum in point x_0. If f(x) is a differentiable function with an existing derivative function f'(x) and at point $x_0 \in I \subset R$ exists $f'(x_0) = 0$, then f(x) has a local maximum or a local minimum in point x_0. Rolle's Theorem states that there exists at least one position $x_0 \in (a, b)$ having $f'(x_0) = 0$. So, this theorem ensures that there exists one local maximum or local minimum at least between points a and b. Consult any calculus book for additional details on Rolle's Theorem.

2.2 Identification of the Peak and Valleys in Time Series Data

The Peak-Valley algorithm uses a geometrical approach to find local peaks and local valleys in a time series. This algorithm detects all local peaks and valleys given in a time series. In this algorithm, a peak cannot be a valley and vice versa. If the points are not a peak or a valley, the algorithm ignores these time series.

Let's take a time series T with n observations. We assume there exists a peak function and a valley function that will identify the local peaks and local valleys. The peak and the valley functions produce values $x_i \in R$ with $i = 1, \ldots, n \in N$. A peak function S produces for a local peak a positive value. We can define the set of local peaks as P with P= $\{(t_i, x_i)|S(x_i) > 0\}$ with $i = 1, \ldots, L$. The valley function is vice versa to the peak function. ` 0 local valleys V= $\{(t_i, x_i)|S(x_i) \leq 0\}$ with $i = 1, \ldots, L'$. Then, from these points, we can estimate the mean, variance, and the standard deviation. See Schneider (2011) for the analytical details on the latter and how the estimation of the peak function is used to determine the peaks in the time series data.

The peak function requires the use of at least three points to determine the peak by comparing the different points. If there are additional peak functions to be used, these peak functions are estimated using the entropy function. Schneider (2011) lays out a nice analytical discussion of the entropy function and uses kernel functions to compute the probability of the values inside the sequence of a given time series to determine if there is a peak or valley in the time series. The entropy of a given sequence is the measurement of disorder in this sequence. Also, the calculation of entropy is based on the probability of the appearance of the values inside the given sequence of the time series, and the entropy function would be minimized. Then, the use of the kernel functions is used to estimate these probabilities to determine the peak and valley in the time series. The general estimate of the kernel function is given as

$$f(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)$$

where h is a function of the number of elements inside the sequence of the time series surrounding a peak or valley. The kernel function K(x) can be replaced by other kernel functions such as the Epanechnikov and the Gaussian. See Schneider (2011) for the analytical details for the calculation of the probabilities of this sequence.

2.3 Methodology in this Paper

The general methodology used in this paper is outlined as follows:

Step 1. Derive a functional relationship between the GDP and time.

Step 2. Derive a functional relation between unemployment and time.

Step 3. Derive a functional relation between inflation and time.

Step 4. Obtain the plot for leading indicators for all the variables of interest.

Step 5. Obtain the plot for lagging indicators which is the unemployment variable.

Step 6. Obtain the plot for coincident indicators, which is the GDP.

Step 7. Obtain the functional relationships for leading, lagging, and coincidental indicators.

Step 8. Apply the algorithm proposed by Schneider (2011) to obtain the local minimum, maximum, and global minimum and maximum based on the content from step 7. This includes the calculation of the minimization of the entropy function as briefly discussed in the previous subsection[1] and the calculation of the probabilities of the points surrounding the potential peak and valley using a kernel function.

3. Empirical Results[2]

This section shows each of the steps in the application of this algorithm to determine peaks and valleys in the economic series.

Step 1 Shows the plot of GDP and time for a linear and nonlinear functions.

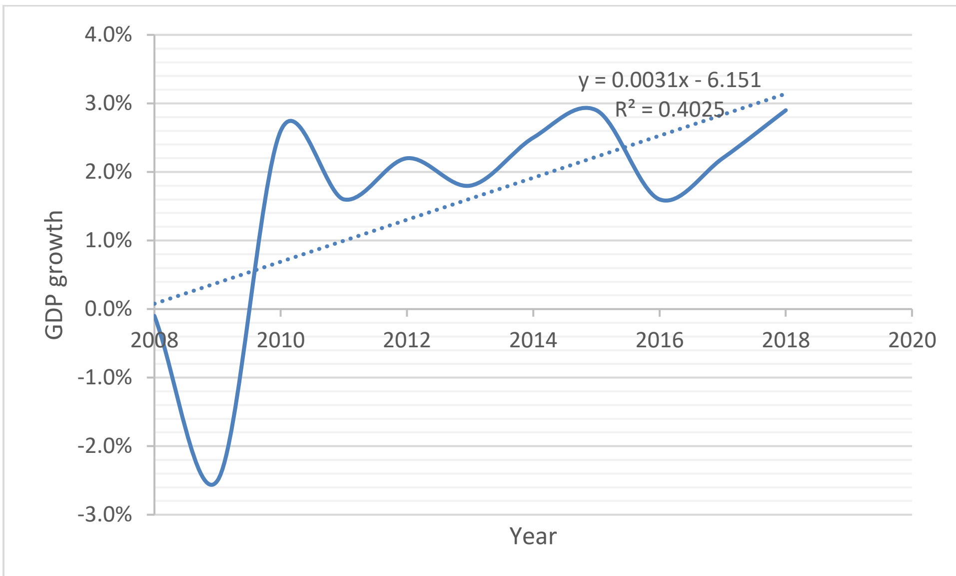

Figure 1: Linear Relation of GDP Growth and Year

[1] The calculations were performed using algorithms in Matlab.
[2] The intermediate output for the application of this algorithm is numerous and is not provided here to maintain brevity. If the interested reader wishes to receive intermediate outputs, please contact the authors.

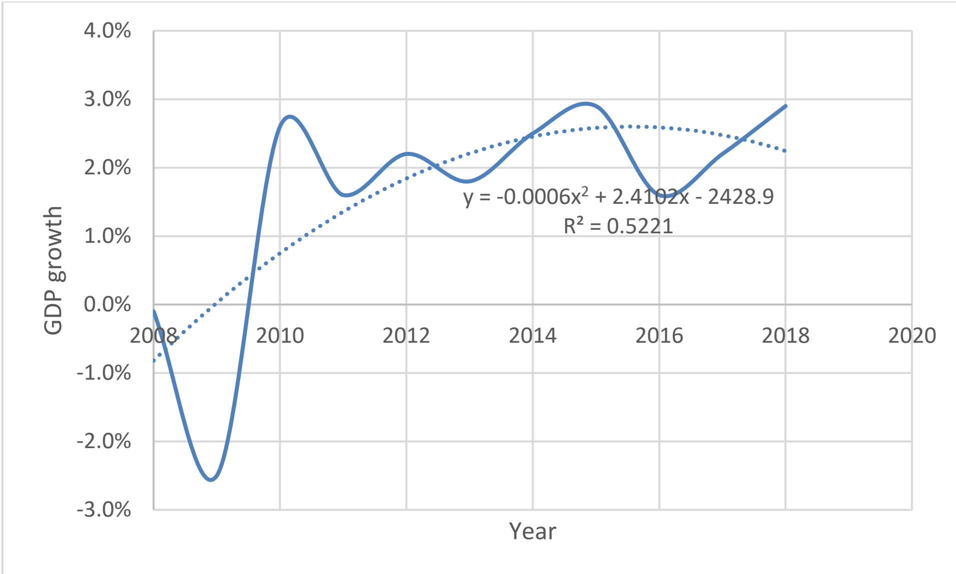

Figure 2: A linear and polynomial of GDP Growth and Year

Step 2—shows the linear and nonlinear relationships of unemployment and year.

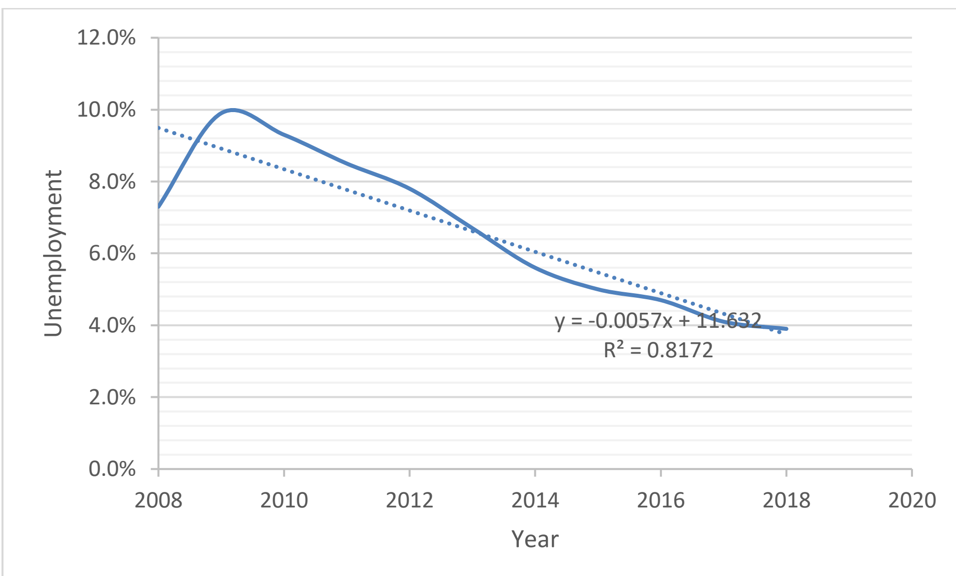

Figure 3: Linear Relationship between Unemployment and Year

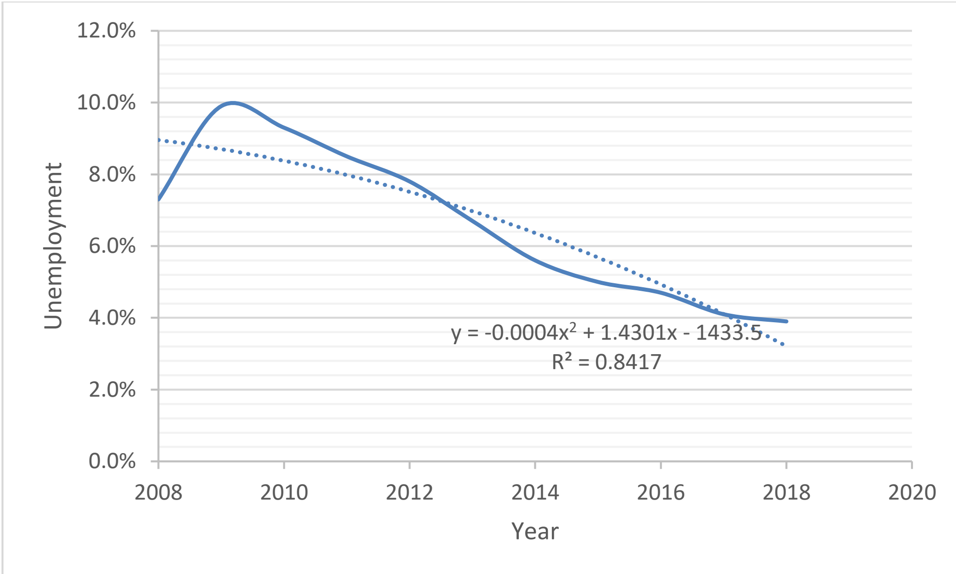

Figure 4: Nonlinear Relationship between Unemployment and Year

Step 3—The graphical depiction of the linear and nonlinear relationships of inflation and year.

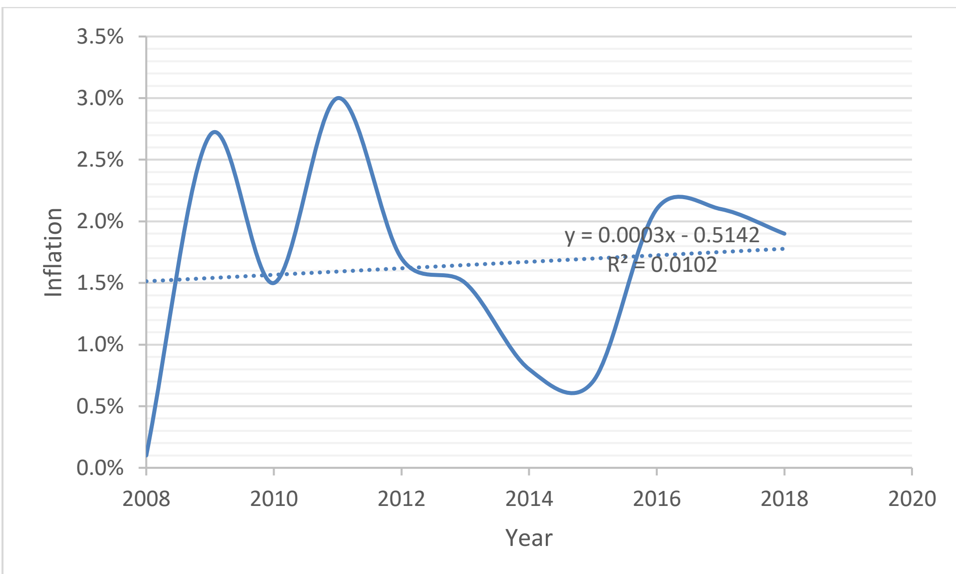

Figure 5: A Linear Relationship between Inflation and Year

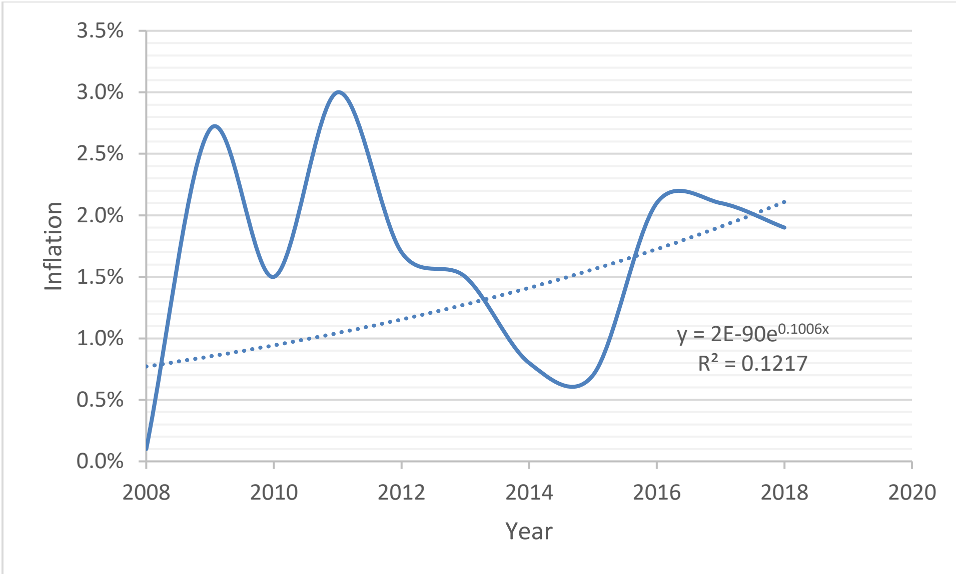

Figure 6: A Nonlinear Relationship between Inflation and Year

Steps 4- 7 Obtained the data from the leading, lagging, and coincident indicators from the Conference Board. These data are summarized in Table 1.

Table 1 Data for Leading, Lagging, and Coincident Indicators from Conference Board[3]

Year	Coincident Indicators	Leading Indicators	Lagging Indicators
1970	62	82	94
1972	60	71	101
1974	63	76	92
1976	67	85	93
1978	70	83	95
1980	71	78	106
1982	71	82	97
1984	72	90	96
1986	75	93	98
1988	82	95	102
1990	84	97	104
1992	85	94	98
1994	92	96	95
1996	95	98	94
1998	104	97	100

[3] We did not present all the time series from the coincident, leading, and lagging indicators as a continuous time series to save on space.

2000	110	110	105
2002	111	109	108
2004	110	111	110
2006	110	113	108
2008	115	115	105
2010	117	117	98

The following figure shows the graph of the information from the preceding table.

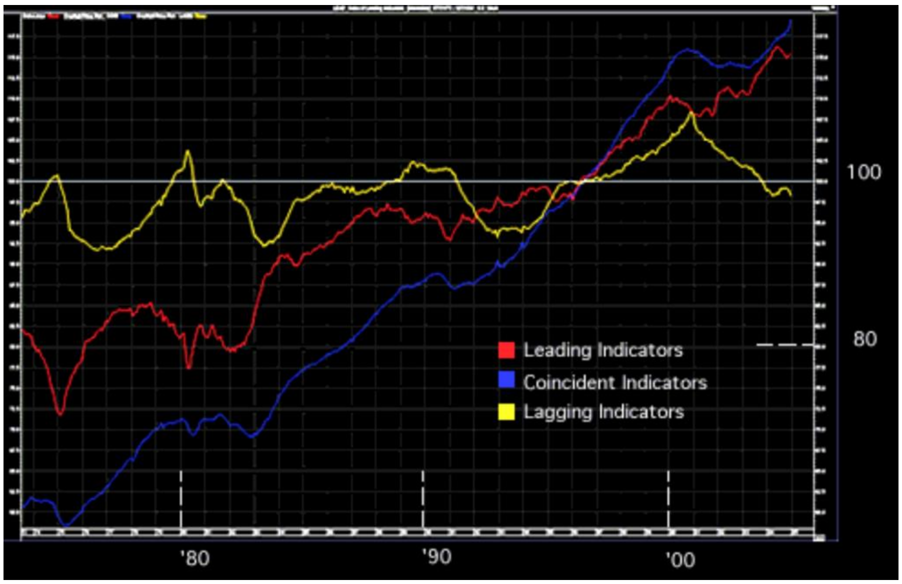

Figure 7: Behavior of Leading, Lagging, and Coincident Indicators

The time series plot of leading, lagging, and coincident indicators is shown in Figure 7. After the plot of these time series, the algorithm by Schneider (2011) is performed to determine the peaks and valleys in the time series data by calculating the minimization of the entropy function.

Steps 5-8 In these steps, we analyze how the variable is behaving over time by looking at the number of local maximums and local minimums as well as the global maximums and the global minimums using the algorithm proposed by Schneider (2011).

The determination of the local maximums and minimums and the global maximums and minimums is calculated the algorithm. These results by leading, lagging, and coincident indicators are presented. In Tables 2 through 4, the time value pair is the result of the entropy function that is minimized and the calculation of the probabilities around the associated points surrounding the eventual peak and valley.

Table 2: Analysis of the Time Series Plot of Lagging Indicators from Figure 7

Extremum	Time Value Pair
Local Peak	(1973,100), (1980,105), (1990,101), (2000,110)
Global Peak	(2000,110)

Local Valley	(1975,95), (1984,96), (1994,94), (2005,99)
Global Valley	(1994,94)

Note: In the parentheses, the years of the local and global peaks are identified along with their values. The latter would also be applied to the global and local valleys.

Table 3: Analysis of the Time Series Plot of Coincident Indicators from Figure 7

Extremum	Time Value Pair
Local Peak	(1970,73), (1980,75), (1990,83), (2000,112)
Global Peak	(2000,112)
Local Valley	(1976,70), (1980,75), (1992,84), (2004,108)
Global Valley	(1976,70)

Note: In the parentheses, the years of the local and global peaks are identified along with their values. The latter would also be applied to the global and local valleys.

Table 4: Analysis of the Time series Plot of Leading Indicators from Figure 7

Extremum	Time Value Pair
Local Peak	(1973,100), (1980,105), (1990,101), (2000,110)
Global Peak	(2000,110)
Local Valley	(1974,76), (1981,79), (1992,95), (2002,105)
Global Valley	(1974,76)

Note: In the parentheses, the years of the local and global peaks are identified along with their values. The latter would also be applied to the global and local valleys.

4. **Discussion of the Results and Conclusions**

The behavior of GDP, unemployment, and inflation is plotted, and the functional relationships were estimated using ordinary least squares (OLS). In addition, the components of the Business Cycle Index (BCI) - Leading indicators, Lagging Indicators and Coincident Indicators - have been studied in detail using the corresponding time series in the literature utilizing the algorithm proposed by Schneider (2011).

From the results, the coincident index gives the lowest mean value of global peaks and valleys which indicates less undulations and hence better measures of the aggregate economic activity. Further study of these data can be performed using the principles of probability and statistics in terms of the exceedance of normal economic limits for leading, lagging, and coincident indicators. That is, to determine the peak and valley, we need at least three points in a range, and any points that are peaks and valleys need to be estimated by the entropy function which is minimized. The probabilities are calculated using kernel functions. The values presented in Tables 2, 3, and 4 show the outcomes from the minimization of the entropy function associated with the probabilities estimated by the kernel functions. In future research, we would like to look at additional economic variables and examine the behavior of those variables. Such variables would include the interest rates, industrial production, and other economic variables. More importantly, we would apply this algorithm to the quarterly data because we used annual data in this paper. Once all economic variables have been reviewed, we can create a CCI that will make it easier to tell us the current state of the US economy.

Acknowledgements

The authors acknowledge the research assistance of Yassine Bouzoubaa at the Department of Civil and Environmental Engineering, California State University, Fullerton and Abdulrahman Saaty at the Department of Civil and Environmental Engineering California State University, Fullerton.

References

Ang, A. H., & Tang, W. (2007). *Probability concepts in Engineering: Emphasis on Civil and Environmental Engineering*, New York: John Wiley and Sons.

Bilana, Y., Gavurovab. B., Stanistawc, G., & Tkacovad, A. (2017). The Composite Coincident Indicator for Business Cycles. *Acta Polytechica Hungarica, 14*(7).

Chapra, S. C., & Canale, R. (2015). *Numerical Analysis for Engineers and Scientists*, New York: McGraw-Hill.

Putcha, C., & Sloboda, B. W. (2017). A Composite Economic Index for the United States' Economy: What Does This Index Tell Us? Presented at the Joint Statistical Meetings, Baltimore, Maryland.

Putcha, C., & Sloboda, B. W. (2018). The Measurement of the Aggregate Economic Performance for the United States via the Composite Economic Index (CEI). Presented at the Joint Statistical Meetings, Vancouver, British Columbia.

Schneider, R. (2011). Survey of Peaks and Valleys Identification in Time Series, Department of Informatics, University of Zurich. Accessed February 4, 2019 at https://www.ifi.uzh.ch/dam/jcr:ffffffff-96c1-007c-0000-00000c9b1cfd/ReportRSchneider.pdf.