

Research on Economic Periodicity Based on Principal Component-Weighted Distance and Clustering Analysis

Jiayi Yan^{1a} Qian Pu^{1b} Junfei Liu^{1c}

¹Qingdao No.2 Middle School Qingdao, China

Abstract—Based on the knowledge of economics, this paper selects 22 macroeconomic indicators that best reflect the overall economic situation of the United States. After differential, logarithmic and exponential preprocessing of the original data, this paper, based on the power spectral analysis model, adaptively identifies the periodicity of the selected economic indicators, and visualize the results. As a result, it screens out 11 indicators with obvious periodicity. In the process of solving the weighted distance based on principal component analysis, correlation test is first conducted on the selected 11 single indicators of periodicity to obtain Pearson correlation heatmap. Then, the principal components are extracted by selecting the first five principal components as the virtual indicators to represent the monthly economic situation, and calculating the weighted distance value between months for visualization. Finally, we select the results of 36 months' smoothing for analysis, figure out the time intervals with similar economic situation, and verify the conjecture of economic periodicity.

Finally, based on K-MEAN clustering analysis, the economic conditions of 352 months are classified into 3 clusters by using the weighted distance after 36 months' smoothing. From the visualized results, it is found that there are two complete cycles, i.e. red-yellow-blue and red-yellow-blue, which is consistent with the conclusion of principal component analysis model, and proves the existence of economic cycle again.

In conclusion, based on the above PCA weighted distance and clustering analysis, it can be concluded that the economic period is around 176 months, in favor of medium long periodicity theory.

1 Introduction

1.1 Research Background

As the famous journalist and historian Eduardo Galeano remarked, "History never really says goodbye. History says, see you later"[1]. For a long time, investors and market observers have always been doing researches on the past markets as an indicator of the current situation and future prospect. As the old saying goes "we should take history as a mirror", history always provides some references for us. Therefore, we need to use historical data for integrated analysis and calculation to summarize the law of economic development. In addition, we need to compare the current economic development with history, and guide the formulation of economic policies in line with the current situation in combination with the past experience and lessons. In history, the theory of economic cycle came into being.

Economic cycle refers to the periodic fluctuation phenomenon in the process of national economic development due to the fluctuation of various economic

indicators [2]. Such phenomenon is a major feature of macroeconomic operation, which can be used to describe the fluctuation of economic activities under the overall growth trend. Generally speaking, the social productivity increases, the level of household consumption rises, the employment status improves, and the employment and output are at their peaks in economic boom. However, the economy gradually declines and falls between prosperity and depression in economic recession. In economic depression, the national economy is characterized by a sharp decrease in social productivity, a significant reduction in the level of household consumption, mass unemployment in the labor market and employment and output at the bottom. After the depression, the economy gradually picks up and enters the recovery period before the second boom, which is a complete cycle [3].

The discussion on economic periodicity is always a hot topic. Given the fact that different schools have different viewpoints, the main difference focuses on whether the periodicity is endogenous or exogenous [4]. On the deep analysis of this problem, John Keynes, a famous modern British economist, explained in 1936 the inevitable regularity of economic development in his

^a2648717728@qq.com ^bpuq2002@163.com ^c1481239253@qq.com

General Theory of Employment, Interest and Money, i.e. the spiral periodic movement. In Keynes' opinion, capitalist system was not as perfect as what the ancient books described because crisis and cycle cannot be avoided. The main cause of the crisis lies in the lack of effective demand, which is rooted in three fundamental psychological laws [5]. Keynes explained the cause of economic cycle from the psychological perspective.

In 1939, Schumpeter summarized three economic cycles, i.e., long cycle, medium cycle and short cycle. Although these three cycles have different lengths and definitions, there is an inclusive relationship in turn among them. The theory of economic cycle has always been having an important impact on the world, and how does the economic cycle forms have always been the main focus of the contradictions among various schools. Some schools insist that it is exogenous, while others hold that it is endogenous. Keynes pointed out in 1936 that economic development will inevitably form a kind of regularity, that is, the periodic movement of fluctuations. He also stressed that the so-called perfect capitalism does not exist because crisis and cycle coexist. The lack of effective demand may cause economic crisis, and the root cause can be explained by three fundamental psychological laws. Similarly, the formation of economic cycle can also be explained by it.

The theory of real economic cycle holds that the real factors beyond the management system represented by technological shock are the important reasons for the emergence of economic cycle [6]. However, the theory of equilibrium cycle, proposed by Robert E. Lucas Jr., the main representative of the Rational Expectation School, which is called Lucas cycle, holds that the main reason for economic fluctuation lies in the response of producers to price changes [7]. The theory combines time series analysis with rational expectations, achieves the microanalysis of macroeconomic and the long-term and dynamic macroeconomic analysis, and maintains the basic theory of classical economics.

In the research process of economic periodicity, scholars gradually turn their research focus from time domain analysis to time-frequency domain analysis, and extract each frequency and amplitude of economic cycle fluctuation from another perspective, which also provides a basis for us to research the periodicity of economic fluctuations. In addition, we also introduce the idea of artificial intelligence into the research of periodic fluctuations, that is, we regard the data indicators of time points as sample points, obtain the monthly classification based with same or similar economic situation on the distance between sample points, and get the results of periodic changes based on the classification results.

1.2 Overview

In order to research the economic periodicity, we select the United States with rich economic statistics for periodicity identification, aiming to take history as a mirror and find out the periodicity of economic development, so as to predict future development and formulate relevant policies. We obtain 22 macroeconomic

indicators from the statistics provided by US officials with a time span from January 1990 to May 2019, totaling 352 months, to analyze the periodicity of single indicator and the overall situation. In this paper, we mainly conduct the following researches:

(1) Adaptive power spectral analysis: After the initialization operations such as first-order difference, second-order difference, logarithmic and exponential preprocessing of macroeconomic data, we conduct power spectral analysis (PSD) separately, select the best for visualization, and classify by whether the single indicator has periodicity;

(2) With selected periodic indicators, we conduct principal component analysis, and demonstrate correlation coefficient in heatmap; after calculating the eigenvalues and eigenvectors, we select the first five principal components for the weighted calculation of Euclidean distance, so as to measure the similarity between months in economic situation;

(3) We use the PCA weighted distance for K-MEANS clustering analysis, and visualize the results in 3D Euclidean space, so as to prove the existence of periodicity.

2 Adaptive Periodicity Identification Based on PSA

2.1 Selecting a Selection of Macroeconomic Indicators

The data set used in this paper is from the United States, whose official economic statistics is open, accurate and timely. The official economic indicators of the United States can be used as a good data sample to research economic cycle. The main macroeconomic indicators selected in this paper are shown in Table I:

TABLE I. TABLE TYPE STYLES

Classification	Indicators	Time Span
Personal	Total personal income	1959.1-2019.5
	Total personal expenditure	1959.1-2019.5
	Individual income tax	1959.1-2019.5
	Personal disposable income	1959.1-2019.5
	Total amount of personal saving deposits	1959.1-2019.5
Employment	Unemployment rate	1948.1-2019.6
	Non-farm payroll employment	1939.1-2019.6
	Labor participation rate	1948.1-2019.6
Prosperity	GDP	1947.5-2019.5
	CPI	1982.1-2019.6
	University of Michigan's Consumer Confidence Index(CCI)	1977.12-2019.7
Price index	Institute for Supply Management (ISM): PMI	1948.1-2019.6
	Export price index: Year-on-year growth of current month	1990.1-2019.6
	Import price index: Year-on-year growth of current month	1983.9-2019.6
Currency	Consumer price index (CPI)	1913.1-2019.6
	Nominal US dollar index® (nominal USDx)	1973.1-2019.6
	M2: Quarterly	1951.1-2019.6
	Federal fund rate (FFR)	1954.7-2019.7
Market	National debt yield (NDY)	1962.1-2019.7
	Monthly average S&P 500 volatility index (VIX)	1927.12-2019.7
Finance	Government revenue	1980.1-2019.6
	Government expenditure	1980.1-2019.6

As the most important indicator to measure national economic development, GDP can highlight the economic prosperity of a country. Similarly, Purchasing Managers'

Index (PMI) reflects the degree of economic prosperity by reflecting the trend of economic change.

As the two most important indicators to measure the labor market of the Bureau of Labor Statistics, unemployment rate and non-farm payroll employment can better affect the trend of all markets compared with all other indicators. CPI is widely used all over the world, which can be used to analyze market prices. It is often used by governments as the basis for developing price and wage policies. It is also used together with personal income and expenditure to measure inflation.

US dollar, as the most special currency in the world, enables the US economic indicators to have a huge impact on the currency world. In order to reflect the situation of US dollar in the international foreign exchange market in an all-round way, we introduce the US dollar index & reg, which can indirectly reflect the change of the export competitiveness and import cost of the United States. In addition, we also adopt FFR, NDY and other indicators in this paper. Similarly, import and export price index and consumer information index can also provide a powerful basis for policy-making.

In light of the different time lengths of various indicators, we select 352 months from January 1990 to May 2019 for research, and fill in the missing data for mean processing. Based on the data, we conduct adaptive periodicity test based on spectral analysis and construct the clustering analysis model based on the weighted distance of principal component to analyze economic periodicity, compare historical data with the current economic development of the United States to draw a conclusion, summarize the historical experience, and put forward reasonable suggestions for economic development.

2.2 Mathematical Principle of Power Spectral Analysis

The Power spectral analysis (PSA) is a signal processing method based on frequency domain analysis. Generally speaking, signals, especially economic indicators, are time series. For a long time, scholars analyze time series based on time domain. However, frequency domain analysis can display the original time signals on the frequency spectrogram, that is, the original signals are decomposed into a number of sine wave components with different periods and amplitudes to obtain more laws that cannot be discovered in the time domain. Under the efforts of scientists, PSA content continues to expand and the method updates rapidly, from which it can be seen that it is a technology with considerable development prospects.

Definition 2.1 (Power Spectral Density). Suppose $X(t), (-\infty < t < +\infty)$ is a stochastic process with continuous mean square.

$$P = \lim_{T \rightarrow \infty} E \left[\frac{1}{2T} \int_{-T}^T X^2(t) dt \right] \quad (2.1)$$

is the average power of $X(t)$.

$$S_x(\omega) = \lim_{T \rightarrow \infty} \frac{1}{2T} E[|F_x(\omega, T)|^2] \quad (2.2)$$

which is power spectral density of $X(t)$.

The Property of Power Spectral Density (PSD): If $\int_{-\infty}^{\infty} |R_X(\tau)| d\tau < \infty$, then $S_x(\omega)$ is the Fourier transformation of $R_X(\tau)$;

$$\begin{cases} S_x(\omega) = \int_{-\infty}^{\infty} R_X(\tau) e^{-i\omega\tau} d\tau \\ R_X(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_x(\omega) e^{i\omega\tau} d\omega \end{cases} \quad (2.3)$$

$S_x(\omega)$ is non-negative real function of ω ; when is rational function of ω , it can be expressed as follows:

$$S_x(\omega) = \frac{\alpha_{2n}\omega^{2n} + \alpha_{2n-2}\omega^{2n-2} + \dots + \alpha_0}{\omega^{2m} + b_{2m-2}\omega^{2m-2} + \dots + b_0} \quad (2.4)$$

Where α_{2n}, b_{2m} are constant, and $\alpha_{2n} > 0, m > n$, the denominator has no real root. Power spectrum of stochastic sequence: In the stochastic sequence $X(n)$, its correlation function $R_X(m)$ satisfies its power spectral density $S_x(m)$, can be expressed as follows:

$$\begin{cases} S_x(\omega) = \sum_{m=-\infty}^{+\infty} R_X(m) e^{-j\omega m} \\ R_X(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_x(\omega) e^{j\omega m} d\omega \end{cases} \quad (2.5)$$

2.3 Conclusions of Adaptive PSA Model

As for the data of 22 indicators collected, we hope to select those indicators with periodic characteristics through spectral analysis. Before the principal component weighted distance model, we introduce difference, logarithm and exponential methods for the preprocessing of original data. For each indicator, we first apply first-order difference, second-order difference, primary logarithmic transformation, secondary logarithmic transformation, primary exponential transformation and secondary exponential transformation, then carry out PSA, and finally extract one of several methods with best periodicity for periodic evaluation.

The criteria for periodic evaluation are as follows: If the period corresponding to the maximum peak of the power spectral density curve is not longer than half of the total investigation time and more than one year, the data is considered to be periodic, and its period is the period corresponding to the maximum peak, otherwise, the original index is considered not to be periodic. We first carry out PSA on the 22 data collected after pretreatment, and obtain the following results:

Figure 1 shows 11 indicators selected with strong periodicity and their corresponding power spectral curves based on the above methods. These indicators are as follows: Unemployment rate (117.33 months), CPI (16.76 months), non-farm payroll employment (117.33 months), nominal USDX (176 months), monthly average S&P 500 volatility index (VIX) (117.33 months), ISM: PMI: Quarterly (70.4 months), export price index: year-on-year growth of current month (176 months), import price index: year-on-year growth of current month (70.4 months), FFR (10.4 months), NDY (70.4 months) and total personal expenditure (176 months).

From the figure, we can see that the period between indicators may be different through the periodic analysis of single indicator. In the research length of 352 months, we can conclude that there are at least two complete cycles in some indicators, which proves the possibility of the existence of periodicity from the perspective of actual values.

Figure 2 shows the indicators that we deem they are not periodic. The visual display results of these 11 indicators show that there are a lot of monotonous growth trends and disorderly fluctuations. After the difference, logarithm and exponentiate of these 11 indicators, we do not find the existence of obvious periodicity, so we may believe that they do not have periodicity under the criteria of this paper.

After PSA of single indicator, we obtain the respective periodicity. According to spectral analysis method, we

conclude that all waveforms can be regarded as the addition of waves with different frequencies and amplitudes. As we consider the economic periodicity of multiple indicators, it can also be regarded as the addition of multiple indicator waveforms in essence. Therefore, the monotonic indicator can be regarded as the general trend of economy, which will have a great impact on the analysis of multi-indicator economic periodicity, so it will not be considered in subsequent calculation.

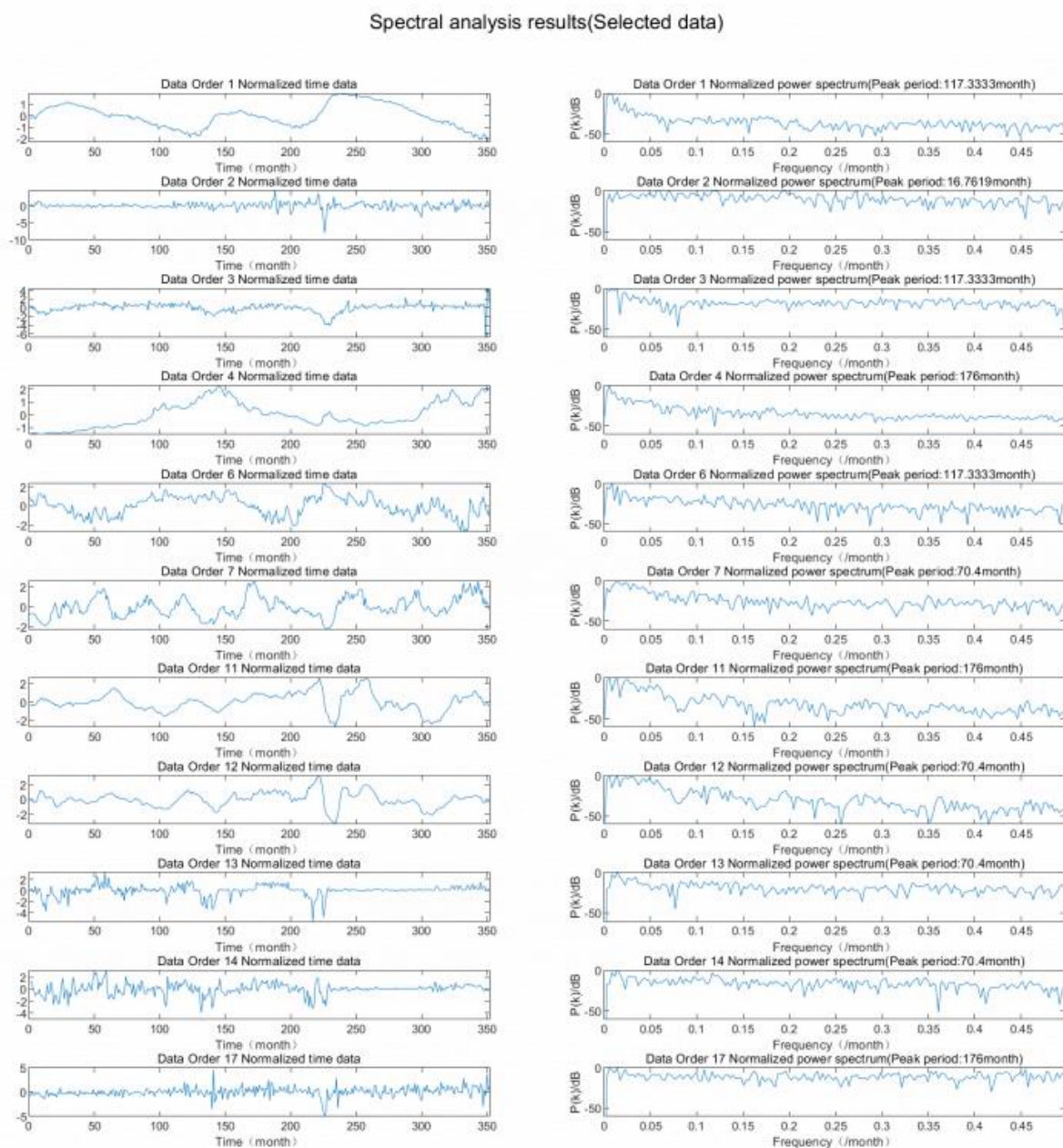


Figure 1. The PSA Results of Periodic Indicators

Spectral analysis results(Unselected data)

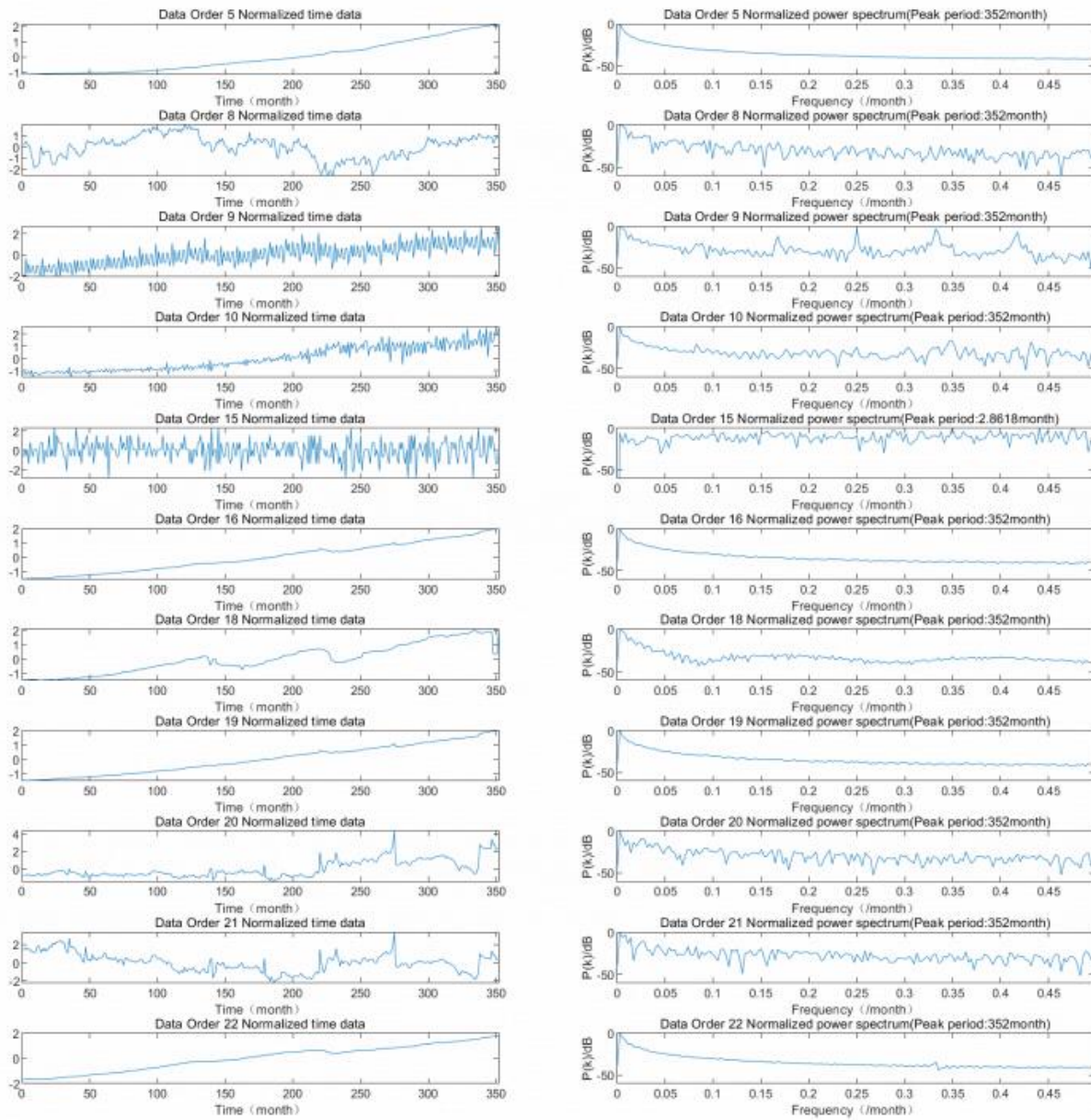


Figure 2. The PSA Results of Non-periodic Indicators

3 Weighted Distance Based on Principal Component Analysis

Principal Component Analysis (PCA) is a multivariate analysis method for non-stochastic variables introduced by Pearson [8] in 1901 in the research process. In 1933, Hotelling [9] further extended this method to stochastic vectors. The use of PCA method for periodic analysis first an reduce the dimension, that is, only a few virtual principal components are selected to express as much information as possible. Under the circumstance where the correlation impact between variables is eliminated, we re-combine the original variables to obtain a new virtual variable, and use the virtual variable for weighted

calculation of the distance between sample points. The advantage of using such multivariate analysis is that it can avoid the repeated calculation of highly correlated indicators, and it is also of practical significance to use eigenvalue of PCA as the weight in distance calculation.

As the indicators to be researched, such as GDP, CPI and unemployment rate, are closely correlated with each other, using expert-defined weight to calculating the distance may result in the impact caused by superposition of different indicators. Therefore, we adopt principal component weighted distance to discuss the similarity of samples.

3.1 PCA Algorithm and Realization

Next, we will conduct Principal Component Analysis to the periodic indicators obtained by Spectral Analysis. The specific realization process is as follows:

(1) Standardized processing of original data: As the indicators are all of specific practical significance, the dimension will have a serious impact on the process of data analysis. In order to eliminate the impact, the data needs to be preprocessed first. In the process of scientific and technological work, commonly used data preprocessing methods include uniformization, standardization and normalization. Here, we use Z-score method to preprocess the data. The calculation of Z-score method is as follows:

$$\tilde{a}_{ij} = \frac{a_{ij} - \mu_j}{s_j} \tag{3.1}$$

Where, $\mu_j = \frac{1}{n} \sum_{i=1}^n a_{ij}$, refers to the sample average of the j indicator; $s_j = \frac{1}{n-1} \sum_{i=1}^n (a_{ij} - \mu_j)^2$ refers to the standard deviation of the j indicator.

(2) Calculation of correlation coefficient matrix: After data standardization, we need to conduct correlation analysis of indicators to obtain the correlation coefficient of different indicators. In this paper, Pearson correlation coefficient is used for correlation analysis.

Definition 3.1 (Pearson correlation coefficient). Pearson correlation coefficient between variables is defined as the quotient of the covariance and its standard deviation between two variables.

$$r_{ij} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \tag{3.2}$$

Where, r_{ij} refers to the correlation coefficient between the i and the j indicators; $cov(X,Y)$ refers to the covariance between variables X and Y ; σ_X and σ_Y refer to the standard deviations of variables.

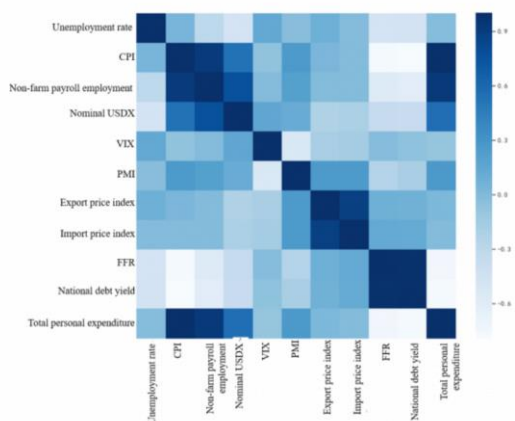


Figure 3. Pearson Heatmap of Indicators

According to Figure 3, it is concluded that there is a strong correlation between nominal US dollar index& reg, non-farm payroll employment & CPI and total personal expenditure, and between import price index and export price index, as well as between national debt yield and federal fund rate. On the one hand, the increase of non-farm payroll employment directly reflects the improvement in employment, on the other hand, it also reflects the potential inflationary pressure. As labor force

plays a vital role in the US economy, the US government will adopt policies to adjust the non-farm payroll employment data to make it in a reasonable fluctuation range. Generally speaking, excess money in circulation leads to inflation. In order to curb the adverse effects of excess, the Federal Reserve will usually raise interest rate to reduce inflation. However, the ups and downs of the US dollar are generally positively correlated with interest rate, and the non-farm payroll employment and total personal expenditure also reflect the economic development of the United States, which directly affects the nominal US dollar index& reg, so there is a strong correlation between non-farm payroll employment and nominal US dollar index& reg. Meanwhile, CPI is also closely correlated with non-farm payroll employment and nominal US dollar index& reg. In sharp contrast to the Chinese people's saving habits, Americans are more inclined to spend the money they earn quickly to maintain at a higher standard of living for the whole family, and they pay more attention to spiritual consumption. In fact, the change of CPI directly reflects the change in residents' purchasing power, that is, the change in total personal expenditure. Therefore, CPI, non-farm payroll employment, nominal US dollar index& reg and total personal expenditure are strongly correlated with each other. FFR and long-term interest rate are separately controlled by the Federal Reserve and influenced by long-term national debt traders, while long-term national debt traders themselves will also make trading decisions according to the policy of the Federal Reserve, which well explains strong correlation between them mentioned before.

As far as the international commodity trade is concerned, on the one hand, the rise and fall of import and export prices of commodities in the world are internally restricted by the scientific and technological content of production technology and labor efficiency. On the other hand, they are externally affected by the change of supply and demand of commodities in the international market and the competition from other countries in the same industry. Therefore, import and export prices are affected by the above factors, and there is a strong correlation between them.

(3) Calculation of eigenvalue and eigenvector:

Definition 3.2 (Eigenvalue and Eigenvector). Assuming that A is an n -order matrix, if there is number λ and vector \vec{v} , let $A\vec{v} = \lambda\vec{v}$, then λ is the eigenvalue of Matrix A , and the vector \vec{v} is the corresponding eigenvector of the eigenvalue λ .

According to the definition, it can be obtained that $(A - \lambda E)\vec{v} = 0$, and the necessary and sufficient condition for the equation to have untrivial solution is: $|A - \lambda E| = 0$. It is a polynomial about λ , which we usually call it eigen polynomial, and we can obtain the corresponding solution of λ by solving the eigen polynomial. After substituting it into the original equation, we can obtain the corresponding eigenvector. The specific calculation results are shown in the table below:

According to Figure 4, we can draw a further conclusion (given the symmetry of the figure, we only discuss the lower triangle area):

(1) First of all, the diagonal in red in the figure refers to the distance between monthly data and itself. In addition, the direction vertical to the straight line can obtain the extension of the blue area, indicating that the economic situation is similar or even identical in several months with certain similar month and even longer time range. We define it as "steady period", which means that there is no obvious change in the economic situation within the period of time. The blue area stretching from the red diagonal becomes wider, indicating that the steady period of economy is longer.

(2) There is another red line approximately parallel with the red diagonal in the figure, where can also find that it corresponds to a dark color zone, which indicates periodic change in economic development, because the line can be deemed that when the horizontal coordinate changes for n months, the vertical coordinate also changes for integral multiple of n , indicating that the month with similar economic development will appear at regular intervals.

(3) Area 1 indicates that the economic situation of two years in early 1990s is relatively close to that from 2007 to early 2011. According to history, in early 1990s, the United States just recovered from Black Monday in 1987; another economic crisis that broke out in 2007 coincided with this, which also proves the periodicity of economic development and the rationality of weighted distance algorithm based on PCA indirectly.

(4) The main purpose of this paper is to identify the circumstance where the current economic environment is closest to history, thus we can conclude that Area 2 in the figure corresponds to the fact that the economic situation from the second half of 2016 to May 2019 is similar to that from the end of 2002 to the middle of 2006. In 2002, the United States just recovered from the economic crisis in 1998. Then in 2006, the US economy was in a stable stage, but in 2007, another serious economic crisis broke out. The upper part of Area 2 and the yellow area at the lower right corner of the figure exactly indicate that the economic environment of that year was not so severe to constitute an economic crisis and damages were caused afterwards. However, combined with the actual situation, our current economic situation is likely to be on the verge of another periodic economic crisis from a historical point of view.

(5) We can see from Area 3 that there are obvious similarities between the economic situation of the United States from early 1992 to late 2000 and that from 2009 to 2018 on a larger time scale. After the great recession in 1991, the US economy began to pick up in 1992, then entered a stage of prosperity in the middle and late 1990s, and has been expanding from 2009 to 2018. The dark blue part of Area 2 is larger than other areas, which indicates that the economic development of the United States has obvious similarity in these two long periods. The dark blue part is surrounded by light blue parts, which indicates that the economic development of the United States has experienced some fluctuations in these two periods, such

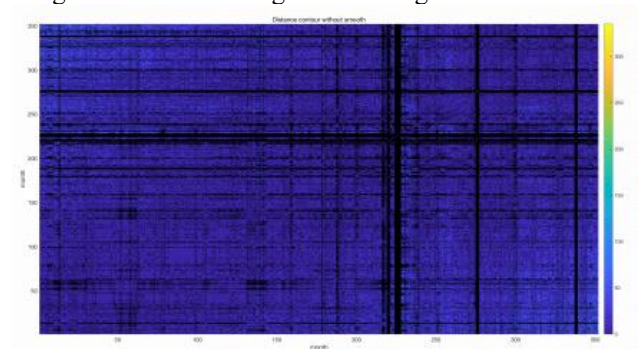
as the twists and turns of the US economy in the first half of the 1990s.

(6) Area 4 shows that the economic situation of the United States from the end of 2002 to the middle of 2006 is similar to that from the middle of 2009 to the early of 2012. From 2002 to 2006, the United States was in economic recovery after the end of the economic crisis, and the economy picked up steadily. Since the outbreak of the global financial crisis in 2008, the United States was shocked and struck in economic recession until the first and second quarters of 2009. It was not until the third and fourth quarters of 2009 that the US economy gradually picked up, but the pace was generally moderate, which is exactly the same as the economic situation from 2002 to 2006. Combined with the conclusion of Area 2, it is not difficult for us to find that the economic conditions of the United States in stages of 2002-2006, 2009-2012 and 2016-2019 are highly similar, i.e. in the recovery state after the economic crisis, which is completely consistent with the conclusion of the model established in this paper. Moreover, these three similar periods of economic development also strongly prove the existence of economic cycle.

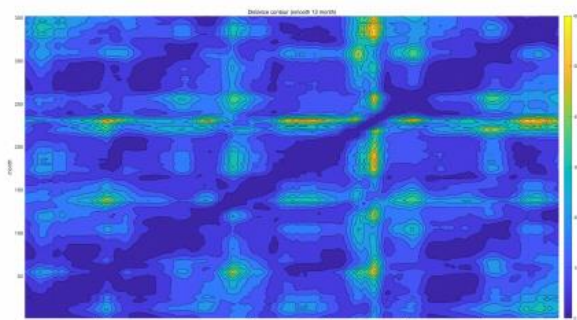
2) Smoothing Analysis of Different Time Scales

In this paper, we conduct the periodicity research based on macroeconomic data. According to the relevant literature, we know that the macroeconomic periodicity is characterized by a long time scale. Therefore, for the monthly data used in this paper, the PCA results will be in high resolution, which is actually not good for us. Therefore, in order to explore periodicity we need to smooth the original data on different time scales.

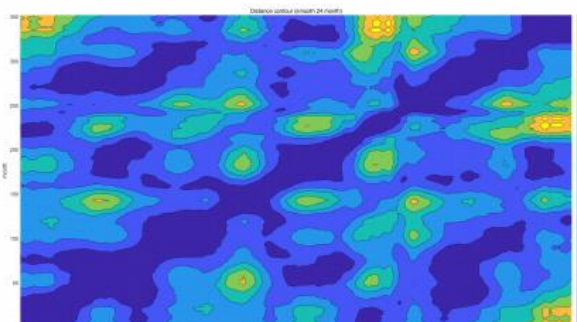
The so-called smoothing is to move and average the data by window size, which can reduce the noise of the original data. As shown in Figure 3.3, the straight line in the figure means that the economic situation of the corresponding month is greatly different from that of all other months. In addition to this, those areas are similar in color. We smooth the PCA data for 12 months, 24 months and 36 months, respectively, and compare it with the image without smoothing to obtain Figure 3.3.



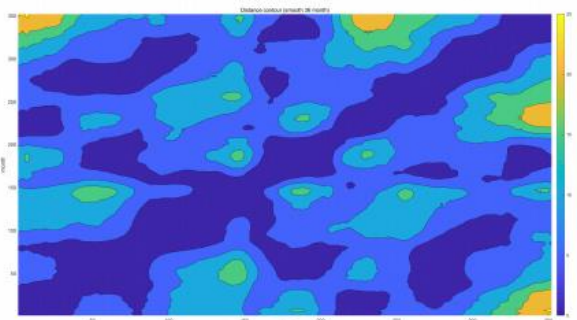
(a) Original Data



(b) After 12 Months' Smoothing



(c) After 24 Months' Smoothing



(d) After 36 Months' Smoothing

Figure 5. Results of PCA Weighted Distances after Smoothing (the contour lines are so dense that the yellow part in Figure (a) is shown in black)

Figure 5: Smoothing Results of PCA Weighted Distances (the contour lines are so dense that the yellow part in Figure (a) is shown in black) According to the results shown in Figure 3.3, when we smooth the original data on the time scales of 12, 24 and 36 months, respectively, we find that the diagonal from the lower-left corner to the upper right corner becomes gradually clear, which indicates that the smoothing does play a role in reducing image noise. In addition, we also obtain the regional characteristics of economic similarity by smoothing the original data, which also indirectly reflects that macroeconomic indicators are characterized by a long cycle.

4 Periodicity Research based on K-MEANS Clustering Analysis

4.1 Mathematical Principle

K-MEANS is known as a classical clustering method. With this method, the data samples can be divided into several categories according to certain similarity indicators. Specifically, "K" in K-MEANS represents the number of categories after clustering. The K-MEANS method is especially for the case where independent variables are continuous values. Simply speaking, K-MEANS divides N samples containing P variables into K categories, making the sum of squares of the distance from internal point to central point of categories minimum. The specific realization steps are as follows:

1. Divide the samples into K categories on the whole, and each sample is randomly divided into one of K categories.
2. Calculate the center of each category. Then, calculate the distance of each sample to the center of each category, select a center closest to it, and divide the sample into the category.
3. Repeat Step 2 until the category of all samples will not be changed, and the algorithm ends.

In K-MEANS method, the distance measurement method is very important. In this paper, we use the weighted distance defined in Equation 3.7 to measure the distance.

Given the large size of samples processed in K-MEANS method and the randomness of initial center generation, the method cannot obtain the global optimal solution. What we obtain is usually the local optimal solution. In this case, any movement of the center cannot make the square of the distance from the internal point to the central point of category smaller, so the clustering reaches a stable state. In practice, we usually run K-MEANS for multiple times, select different starting points each time, and then take the stable state with the best final effect as the final result of clustering.

4.2 Conclusion and Analysis

3) Conclusions of K-MEANS Clustering Model

After smoothing PCA results for 36 months, we conduct K-MEANS clustering analysis and visual display on three-dimensional coordinates. The three dimensions in the figure correspond to the first three principal components, respectively, and red-yellow-blue corresponds to the three categories defined by human, respectively. In this paper, we only use this model to discuss periodicity, and make no difference between the three categories. According to the figure, we can see that the monthly data has gone through two complete cycles of red-yellow-blue and red-yellow-blue, which is consistent with the conclusion of the previous chapter.

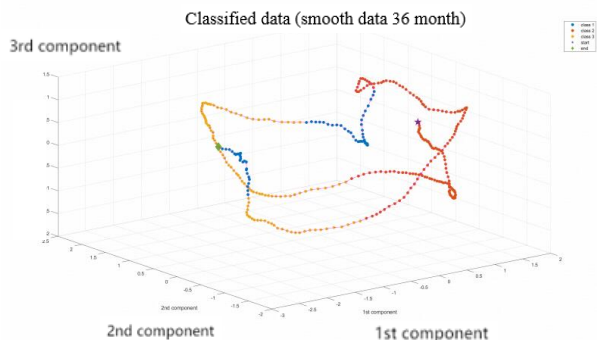
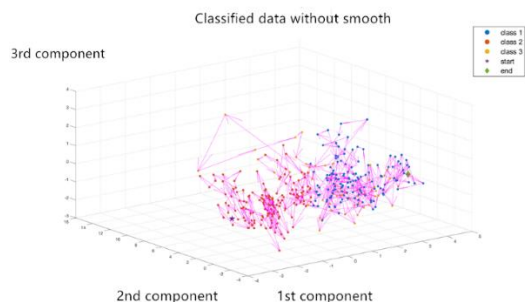


Figure 6. K-MEANS Clustering Analysis Results of 36 Months' Smoothing

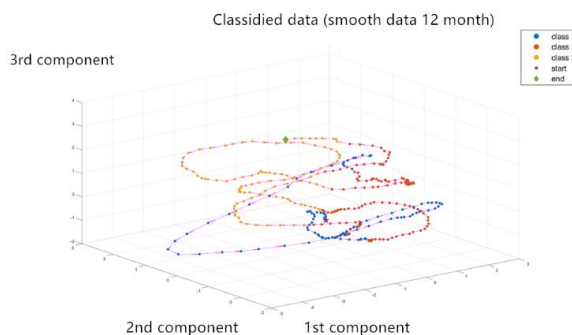
In addition, the two curve sections in the same color are relatively similar in length, which indicates that the stable cycle of economic situation is also basically the same. That is, in each section of same or similar economic situation, the length of time to maintain such economic stability is the same, which also favorably proves the existence of periodicity of 176 months in favor of medium long period.

4) *Smoothing Analysis of Different Time Scales*

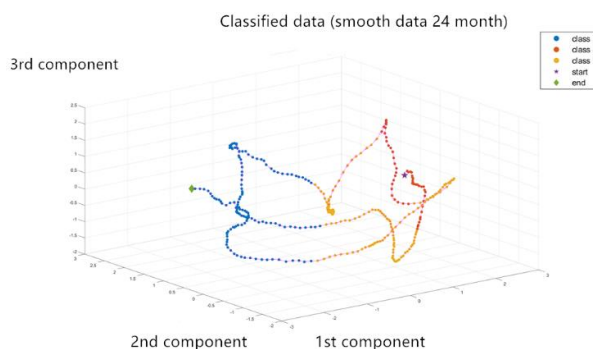
Component Similarly, we still discuss the visualized results after smoothing on different times scales after we obtain the results of K-MEANS clustering analysis. First of all, we can see from the results of original monthly data (without any smoothing) that when the categories are divided into three categories, monthly data can also be divided into three categories by distance, but the data points are relatively chaotic. We can know that the current month is similar to certain month in history and the current economic environment is similar to certain stage in history, which essentially improves the resolution of clustering analysis.



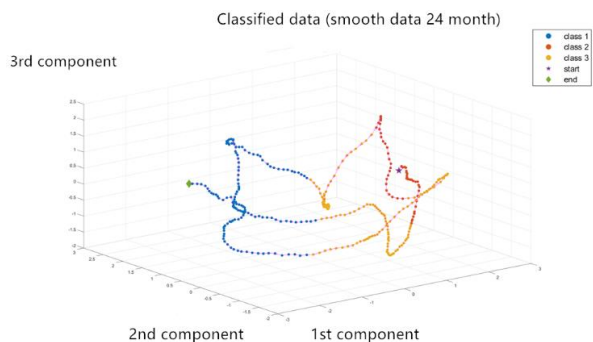
(a) Original Data



(b) After 12 Months 'Smoothing



(c) After 24 Months 'Smoothing



(d) After 36 Months 'Smoothing

Figure 7. K-MEANS Clustering Analysis Results after Smoothing

In addition, when the data is smoothed for 12 months, we can actually see that the economic situation tends to be stable in a long period of time, and undergoes periodic changes among these three categories. But due to the long time scale of economic cycle, we can still see the impact of noise. With the increase of smoothing scale, when the data is smoothed for 36 months, we can see that the economic situation has undergone two complete cycles of red-yellow-blue and red-yellow-blue since January 1990, which is consistent with the conclusion of weighted

distance algorithm based on PCA. In this model, data smoothing does greatly eliminate the impact of noise.

5 Summary and Prospect

5.1 Main Conclusions

In this paper, we have conducted PSA, principal component weighted distance calculation and K-MEANS clustering analysis on 22 economic indicators, and find that the analysis object, namely, the US economy, has obvious periodic characteristics in the period of 352 months from January 1990 to May 2019.

(1) In the adaptive period recognition model based on spectral analysis, we first use difference, logarithm and exponentiate methods to preprocess the original data of the 22 indicators collected, later conduct PSA, and we select the best result as the periodic analysis result of single indicator, which is called adaptive algorithm. After the visual display of the spectral analysis results, we obtain 11 indicators with periodicity. Therefore, those indicators without periodicity will be deleted to avoid causing interference to the analysis of economic periodicity.

(2) We use the weighted distance calculation method based on PCA to obtain the distance between monthly data from January 1990 to May 2019, and obtain the corresponding thermodynamic diagram. As economic periodicity is a research topic based on a longer time scale, we smooth monthly data for 36 months. By analyzing the visualized results, we obtain the regional characteristics that reflect economic similarity, and analyze the historical economic situation of representative areas.

(3) After obtaining the results based on principal component weighted distance, we further conduct K-MEANS clustering analysis according to the distance obtained, and achieve visual display on the three-dimensional coordinates, from which we can see that these 352 months have undergone two complete cycles, which is consistent with the conclusion of PCA. The identified 176 months period is consistent with medium long period theory.

5.2 Prospect

In this paper, we obtain the results of macroeconomic periodicity based on multiple models, and find that our models have some objective shortcomings while identifying where the current economic environment is.

(1) The data length selected in this paper is 352 months. For the research of the economic cycle reflected in the long-term economic situation, the data length is too short. Next, we will use data of a longer period of time as much as possible to research the issue of economic periodicity.

(2) In the spectral analysis of macroeconomic indicators, we can see that high and low-frequency noises have impacts on the results, which also causes some interference to PSA results. Therefore, we shall conduct high-pass and low-pass filtering on original data to improve the resolution of the results.

(3) By analyzing the original indicators, we obtain only 11 indicators that can be used for PCA, and the data dimension is smaller, which may inevitably make the conclusion of economic periodicity biased to some degree. When the data dimension and data coverage are further increased, the data models will play a better representative role.

(4) In the last part of this paper, we adopt a simple method, i.e. K-MEANS, to conduct clustering analysis on the principal component selected, which is more like a supplement to PCA in essence. In the follow-up work, we may use a higher-order algorithm, such as neural network, etc., to identify the mode of economic situation.

Acknowledgment

History never ceases and the traces of history will last long. For a long time, investors and market analysts have always been taking the research on past markets as an indicator to measure the current situation and judge the prospects. In today's high-speed economic development, the economic situation is complex and changeable, but there is always an "invisible hand" behind the economic market to push it forward. Therefore, we suppose that the current economic situation is highly coincident with the economic situation of a certain historical period, and there may be more than one such period. Thus, we assume that there is a certain periodicity in the development of economic situation. When searching for the relevant information of school topic research on the Internet, we found an article on the website of "TWO SIGMA" that weights some representative indicators to discuss and analyze the historical situation of the US financial market, so it reminds us that we may use the relevant knowledge of self-taught machine learning and mathematical modeling to research economic periodicity. In the process of writing this paper, we worked together and completed different parts of the project. Specifically, Yan Jiayi is responsible for the research of spectral analysis and the writing and proofreading of this paper; Pu Qian is responsible for the operation of PCA and the data visualization; Liu Juanfei is responsible for the modeling of K-MEANS, as well as data collection and processing. In the three months' research of economic periodicity, we have benefited a lot. We have learned a lot of knowledge of economics, felt a sense of achievement in mathematical modeling, and established a deep friendship. Whether in classroom or after-school research, we are willing to be each other's teacher and make progress together.

We hereby want to extend the special gratitude to our supervisor—Mr. Ma Lianhan. As this paper is mainly written after class, Mr. Ma sacrificed his valuable rest time to answer our questions, and guided us to use Python for data visualization. When we were stuck in bottleneck, delayed the task of modeling unceasingly and even thought about giving up, Mr. Ma talked with us during self-study at night and encouraged us to move on. Thanks to Mr. Ma's diligent cultivation, we can thrive and keep going on the road of modeling. We solemnly declare that all the content of this paper was completed by three of us independently. Thanks to Mr. Ma Lianhan for his

academic guidance and technical guidance in terms of data visualization.

References

1. Bradley Kay, 2018. Markets in the Rear-View Mirror (May Be Closer Than They Appear).www.twosigma.com.
2. Wang Xinli, 2004. Real Business Cycle Theory: Retrospect and Prospect [C]. Quantitative Economics Research.
3. Yao Min, Zhou Chao, 2013. Research on the Characteristics and Influencing Factors of China's Economic Cycle Fluctuation[J], Inquiry Into Economic Issues.
4. Li Jianwei, 2015. Traditional Business Cycle Theory and Its Limitations[M]. Journal of Chongqing University of Technology (Social Science), 29 (11).
5. Xu Shengyu, Ge Yang, Comparison of Marxist and Keynesian Business Cycle Theories [B]. Economist.
6. Huang Zelin, 2005. China's Business Cycle Characteristics and Fiscal Policy Effects: Empirical Analysis Based on Three-sector RBC Model [J]. Economic Research Journal.
7. Lucas, Robert E, Jr, 1978. Unemployment Policy[M]. American Economic Review, American Economic Association. 68(2):353-357.
8. Pearson, K, 1901. On Lines and Planes of Closest Fit to Systems of Points in Space[M]. Philosophical Magazine 2:559-572.
9. Hotelling, H, 1933. Analysis of a Complex of Statistical Variables into Principal Components[J]. Journal of Educational Psychology. 24:417-441.