Application of empirical mode decomposition to improve deep learning for US GDP data forecasting

Shih-Lin Lin

Graduate Institute of Vehicle Engineering, National Changhua University of Education, No.1, Jin-De Road, Changhua City, 50007, Taiwan

ARTICLE INFO

Keywords:
Econometrics data forecasting
Deep learning
Empirical mode decomposition

ABSTRACT

The application of deep learning methods to construct deep neural networks for the prediction of future econometric trends and econometric data has come to receive a lot of research attention. However, it has been found that the long short-term memory (LSTM) model is unstable and overly complex. It also lacks rules for handling econometric data, which can cause errors in prediction and in the actual data. This paper proposes an empirical mode decomposition (EMD) method designed to improve deep learning for understanding US GDP trends and US GDP data prediction research. The US GDP growth rate is used only for LSTM model prediction and for real data comparison; the root mean squared error (RMSE) is 2.7274. The US GDP growth rate is EMD decomposed to obtain the intrinsic mode functions (IMFs) after which the LSTM model is used to predict an RMSE of 0.93557.

1. Introduction

In recent years, artificial intelligence technology has had a big impact on the global business environment stimulating active investment in this type of research in many different fields. At the international level, everyone expects the application of artificial intelligence at critical moments of great economic change [1, 2, 3, 4, 5, 6]. Artificial intelligence methodologies are based on large amounts of data. In the economic context they can quickly sort, screen and analyze data to help the user make more precise investment decisions, significantly reduce human error, bring better returns to customers, more accurately predict possible outcomes and facilitate risk control.

Emerging econometrics analysis methods include estimating the covariance and variance of prices and rewards [7]. Examples of such models include the autoregressive conditional heteroskedasticity (ARCH) model developed by Engle [8], the generalized (G)ARCH model developed by Bollerslev [9], and the stochastic volatility (SV) model of Taylor and Harvey et al. [10, 11]. Notably, the 2003 and 2011 Nobel Prizes in Economics were awarded to American economists Engle et al., who proposed ARCH models and vector autoregressive (VAR) models for time series prediction [8]. Institutional investors and individual investors are struggling to find better investment strategies in anticipation of profitability in the securities market. Profitability is the goal that every investor pursues, but it is not easy to attain. Furthermore, having to pay trading commissions mean that same price trading produces a loss. Finding an effective prediction model to improve prediction accuracy to act as a reference for investors’ investment decisions would help to overcome these problems but it remains a challenging topic for the industry and academia. Bollerslev, Engle, and Kroner proposed a multivariate GARCH model [12, 13, 14] for this purpose. Progress in econometrics analysis can be seen in the chronology of major events outlined below. American economist Markowitz introduced the portfolio theory in 1952 [15]. In order to accurately predict the characteristics of a econometric time series, it is necessary to establish a prediction model that is consistent with its characteristics. It is important to study trends and develop more accurate models. The application of deep learning to time series problems with nonlinear features is an effective method for this purpose [16]. Deep neural networks have already been used to predict trends in future engineering data, but deep learning models have been found to produce errors when used for the prediction of economic data based on actual data.

In response to such problems, the empirical mode decomposition method, the Hilbert-Huang Transform, which is a non-stationary data analysis method, was developed. Any non-stationary data are first subjected to a sifting process to produce IMFs, which comprise the original dataset of non-stationary data. The IMF is used as the basis of expansion. It can be nonlinear or non-stationary. Based on the nature of the original data, all IMFs can be added together to restore the original dataset. In
1998, Huang proposed the Hilbert-Huang transform (HHT), which has since received extensive attention from the academic community. It is an effective method for analyzing nonlinear non-stationary time series, including the empirical mode decomposition method and Hilbert transform [17, 18].

The empirical mode decomposition method has found applications in many different fields [19, 20]. Lin and Huang proposed an improve deep learning method for accurate prediction of the Taiwan CSR Index [21]. The CSR data [21] is Taiwan’s data is local financial data. This research focuses on the development of the US economy because the data on the US GDP growth rate is economic data. Economic data and financial data are not the same theoretical basis. In this study, the US GDP growth rate data is international, and it is more difficult to accurately predict, and it has made a greater contribution to the development of the international economy. Huang et al. used empirical mode decomposition (EMD) to analyze the variability of the economic market and were the first to apply EMD to economic data. They emphasized that the IMF components can be used for filtering. Compared with the wavelet transform and Fourier transform methods used at the same time, it is believed that HHT provides better frequency analysis [22].

Huang et al. argued that the goal of EMD disassembly is to obtain an IMF with single component characteristics, which means that the IMFs must be orthogonal. However, in the actual decomposition process, there is no guarantee that each IMF will be completely orthogonal. In order to ensure a better IMF, the orthogonal index, the confidence limit and the confidence interval are defined [23]. Wu and Huang revised the original EMD and proposed an improved method called “Ensemble EMD” (EEMD), which can greatly reduce on the possibility of mode mixing [24]. Many improved EMD methods have been developed, for theoretical or practical applications, based on their research [25, 26, 27, 28, 29, 30].

This study uses the original EMD because of the more extensive parameter selection required for the modified EMD. For example, although the EEMD can improve mode mixing and ending effects, adding white noise to the signal, the noise parameters are based on user case studies. If the noise parameter is too small or too large, the original signal will be distorted. Therefore, the basic original EMD method is used here.

The US GDP data are broadband in the frequency domain, containing a wide range of fluctuations, so it is not sufficient to use long short-term memory (LSTM) methods for prediction. This study uses EMD to decompose the original non-linear and non-cyclical US GDP data into a limited number of IMFs and trends. The bandwidth of these IMFs becomes narrower and has regular cyclic, periodic or seasonal components in the time domain. LSTM is a good way to predict periodicity or seasonality. All the IMFs are then added together to get the prediction result. Compared with LSTM alone, the EMD plus LSTM model can reduce the root mean square error (RMSE) compared with the actual data. This study improves the accuracy of the prediction model and then describes the application of deep learning and EMD for the prediction of US GDP data. The global economy was greatly affected by the COVID-19 epidemic in 2020, and the US GDP is an important indicator of its impact. Everyone is deeply interested in understanding future trends in the US GDP. This paper is helpful to obtain a deeper understanding of the application of deep learning in economic data forecasting and the method described therein greatly improves the accuracy of the forecasting model for the US GDP.

2. Description of the research data

The GDP represents the final outcome of a country’s production activities (including products and services) over a certain period of time. It is the sum of the added value of various sectors of the national economy during the accounting period. It can provide a complete picture of a country’s economic situation. Governments judge whether the economy is shrinking or expanding, whether it needs stimulation or control, whether it is in recession or under threat of inflation, based on the GDP. It can also reflect the poverty and wealth of a country and the average living standards of the people. Therefore, all governments have made raising the growth rate of the GDP a top priority. The United States is the world’s number one economic power. In 2002, 2003 and 2004, the US economy accounted for 29.6%, 33.8%, and 34.5% of the global economy, respectively. Its contribution to global economic growth was as high as 32.3%, 30.9%, and 34.5%, respectively. Being so influential worldwide, many scholars are studying the future development of the US GDP. The commonly used source for growth rate data is https://fred.stlouisfed.org (30 December 2020), starting with 282 datasets from 01-01-1948 to 04-01-2018. This includes four datasets for each year divided according to seasons, each of which are three months long. All the data are divided into two parts, part of which is used for training (90% of the total data or 254 datasets, from 01-01-1948 to 04-01-2011). The other part is used for verification (10% of all data for a total of 28 datasets, from 07-01-2011 to 04-01-2018).

3. Results and discussion

A brief explanation of the US GDP growth rate is given below. The time frame is from 01-01-1948 to 04-01-2018. The statistical characteristics for a total of 282 data sets are shown in Figure 1. After the Second World War, the economic strength of the United States rapidly
Figure 3. RMSE results from the LSTM model for the US GDP growth rate forecast and actual verification data. The RMSE is 2.7274.

Figure 4. EMD decomposition is the decomposition of the US GDP growth rate into the combination of IMFs, (A) IMF1, (B) IMF2, (C) IMF3, (D) IMF4, (E) IMF5, (F) residue.
increased, giving them a comprehensive advantage in the capitalist world economy. Since the completion of the transition from a wartime economy to a peacetime one in the 1950s, the US economy has continued to grow, maintaining its dominant position. From 1955 to 1968, the gross national product of the United States grew at a rate of 4% per year. From the 1950s to the 1960s, economic growth appeared to create a so-called "golden age", according to Western economists. The growth of the US gross national product in that "golden age" rose from $523.3 billion in 1961 to $106.34 billion in 1971. Industrial production in the United States increased at an annual rate of 18% from 1965 to 1970. In 1970, the United States had 25% of the world's coal production, 21% of its crude oil production, and 25% of steel production. In 1971, the population of the United States owned 111 million vehicles, with 83% of households having at least one car. After the war, the first economic recession was experienced, from 1948 to 1949, during which industrial production fell by 8.3% and the unemployment rate reached 5%. As can be seen in Figure 1, the third quarter GDP growth rate in 1949 fell to a low of -1.5. In the second recession which lasted from August 1953 to August 1954, industrial production fell by 9.1%, and
the unemployment rate reached 6.2%. As indicated in Figure 1, the second quarter GDP growth rate of the first quarter was -2.4. The third economic recession was from July 1957 to April 1958. Although short in time, it was serious in degree. Industrial production plummeted by 13.5%, and the unemployment rate rose to 7.5% in 1958 (see Figure 1). The quarterly GDP growth rate was low, -2.9. From February 1960 to February 1961, the United States experienced its fourth post-war recession. Industrial production fell by 8.6% with an unemployment rate of around 7%. The 1960 fourth quarter GDP growth rate fell to a low of -0.7. From October 1969 to November 1970, the United States experienced the fifth post-war recession. As can be seen in Figure 1, the 1970 third quarter GDP growth rate was -0.2. Then there was the oil crisis from 1973-1975 and, in October 1973, the fourth Middle East war broke out. The Arab oil-producing countries cut oil output, causing oil prices to rise, immediately disrupting the pace of economic development in Western countries, which triggered an economic crisis. In December 1974, one year after the crisis, there was a fall in production in the US auto industry by as much as 32%, and the Dow Jones stock price average index fell to nearly half of the highest point before the crisis. In 1975, the unemployment rate in the United States rose to 9.2%. The GDP growth rate in the fourth quarter of 1974 was -2.3. On October 19, 1987, the US Dow Jones index fell sharply, plunging 508.32 points in one day, down 22.6%, with the highest points and magnitudes of declines before this. In July 1990, the US economy entered a brief, moderate recession when Iraq’s invasion of Kuwait on August 2 caused international oil prices to soar. The fourth quarter GDP growth rate in 1990 was -1.0. During the collapse of the Internet bubble in 2000, the United States experienced a rapid economic recession, including the collapse of some manufacturers (the most prominent example being the energy company, Enron), rising unemployment, shrinking consumption, and overcapacity. The subprime mortgage crisis from 2007 to 2008 was caused by a sudden increase in the default rate and credit crunch in the US subprime mortgage industry, and the shocks, panic, and crisis in the international economic market that
began in the summer of 2007. The 2009 first quarter GDP growth rate was -3.9. The average GDP growth rate in the statistical data is 3.3007, the standard deviation is 2.5913, the variance is 6.7151, for a total of 282 data.

The US GDP growth rate data start from 01-01-1948 and go to 04-01-2018 with a total of 282 data. The data are divided into two parts with 90% used for training (a total of 254; the start time is from 01-01-1948 to 04-01-2011), and the other 10% are used for verification (28 data, starting from 07-01-2011 to 04-01-2018).

This study uses LSTM regression networks for GDP growth rate forecasting. The parameter settings for the LSTM model are as follows. The specified LSTM layer has 200 hidden units. First, the adaptive moment estimation optimizer is selected and trained for 250 periods. To prevent system divergence, the gradient threshold is set to 1. An initial learning rate of 0.005 is specified and the learning rate is reduced by multiplying it by a factor of 0.2 after 125 epochs.

The following formula is used to calculate the root mean square error (RMSE) for verification:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}$$

where \(\hat{y}_i\) is the real data (verification); \(y_i\) is the prediction data. A small RMSE means that the prediction is close to the real data, and the larger the RMSE, the greater the difference between the predicted data and the real data.

First, the adaptive moment estimation optimizer is selected and trained 250 times. To prevent a gradual explosion, the gradient threshold is set to 1. An initial learning rate of 0.005 is specified and the learning rate is reduced by multiplying it by a factor of 0.2 after 125 iterations. Figure 2 shows the results of the LSTM’s GDP growth rate forecast. Figure 3 shows the LSTM’s US GDP growth rate forecast and actual data verification results. The RMSE is 2.7274.

The US GDP growth rate from 2011 to 2018 is stable, fluctuating between 2-4. However, the LSTM prediction result has a large variability period, with the lowest point being close to -3 in the middle of 2012, and...
the highest point being more than 7 in the middle of 2015. It is a coincidence that the US GDP growth rate was 1.383 in the second quarter of 2016 and 2.865 in the first quarter of 2018. The LSTM accurately predicted the same result.

There are two perspectives in traditional time series analysis modeling, linearity and nonlinearity. Intrinsic features are usually smoothed out during the analysis process, so the fluctuation governing the sequence cannot be completely studied. The application of the EMD decomposition method for nonlinear, non-stationary sequences has great advantages. The method allows analysis of the frequency domain, with the original data sequence decomposed into sequences containing different frequencies. The decomposed data can reflect fluctuation information on different time scales while retaining the characteristics of the original data. EMD decomposes the data to be analyzed into a number of components that are locally symmetric to zero average value, which is called IMFs. The empirical mode decomposition is to decompose the signal into a combination of IMFs, and each IMF meets the two conditions mentioned above. Here is the definition of IMF [17]. Any piece of data that satisfies the following two conditions can be called an IMF.

The sum of the number of local maxima and local minima must be equal to the number of zero crossings or can only differ by one at most.
At any point in time, the upper envelope defined by the local maximum and the lower envelope defined by the local minimum should be averaged close to zero.

Therefore, if a piece of data belongs to IMF, it means that its waveform is locally symmetrical to zero average value. EMD is to find out IMF step by step through the continuous repetitive screening process (Shifting Process). The process is as follows.

**Figure 15.** LSTM prediction and actual verification results for IMF4; RMSE of 0.42425.

**Figure 16.** EMD analysis of the US GDP growth rate to get IMF5.

**Figure 17.** LSTM prediction results for IMF5.
I. Find all extreme points of the original signal.

II. Use the cubic spline to find the upper envelope defined by the local maximum and the lower envelope defined by the local minimum.

III. Calculate the mean value of the envelope.

IV. Take out the components.

V. Repeat steps (I)-(IV) until the component meets the definition of IMF.

VI. Calculate the residual quantity (residue).

VII. If the residue is a trend component, the algorithm stops; otherwise, repeat steps (I)-(IV) to find other IMFs.

The SD of the two consecutive screening results is used as the stopping criterion. The SD parameter is selected here as 0.2. When the SD value is less than the given threshold, the screening action will be stopped. Figure 4 shows that the EMD decomposition is the decomposition of the US GDP growth rate into the combination of IMFs, (A) IMF1, (B) IMF2, (C) IMF3 (D) IMF4 (E) IMF5 (F) residue. The GDP growth rate is first decomposed into short, medium and long-term time series components. Here, there are six components: IMF1 to IMF6 (residue). In terms of statistical characteristics, this is a violent high-frequency set of data with an average period of 10.1566 months. The mean is 0.0169, the standard deviation is 0.7393, the variance is 0.5466, and the Pearson correlation coefficient is 0.3510. IMF1 is a high-frequency dataset. It can also be seen that there is a difference after 1983. Prior to 1983 the amplitude of the change is large, but after 1982 the amplitude of the variation is small.

Figure 5 shows the LSTM prediction results for IMF1, and Figure 6 shows the LSTM prediction and actual verification results for IMF1 with an RMSE of 2.7274. The IMF1 and the LSTM predicted waveform results are similar, the highest point is 0.5, the lowest point is -0.5, but the LSTM prediction result is -0.8. The US GDP growth rate has 9 cycles, but there are only 3 cycles in the LSTM prediction results. Figure 7 shows the results of the EMD analysis used to get IMF2. Its statistical characteristics are as follows: the average period is 24.7941 months, the mean is 0.0231, the standard deviation is 1.5956, the variance is 2.5459, and the Pearson correlation coefficient is 0.5923. This is the second highest-frequency dataset. The difference between the period after 1985 and the preceding period is significant.

Figure 8 shows the results of the EMD analysis used to get IMF3. Its statistical characteristics are as follows: the average period is 36.1827 months, the mean is 0.0341, the standard deviation is 2.1568, the variance is 4.6396, and the Pearson correlation coefficient is 0.7814. This is the third highest-frequency dataset. The difference between the period after 1985 and the preceding period is significant.

Figure 9 shows the results of the EMD analysis used to get IMF4. Its statistical characteristics are as follows: the average period is 72.3654 months, the mean is 0.0681, the standard deviation is 4.3137, the variance is 18.6293, and the Pearson correlation coefficient is 0.8375. This is the fourth highest-frequency dataset. The difference between the period after 1985 and the preceding period is significant.

Figure 10 shows the results of the EMD analysis used to get IMF5. Its statistical characteristics are as follows: the average period is 144.7307 months, the mean is 0.1362, the standard deviation is 8.6273, the variance is 74.0758, and the Pearson correlation coefficient is 0.8375. This is the fifth highest-frequency dataset. The difference between the period after 1985 and the preceding period is significant.

Figure 11 shows the results of the EMD analysis used to get IMF6. Its statistical characteristics are as follows: the average period is 289.4613 months, the mean is 0.2724, the standard deviation is 17.3140, the variance is 300.4931, and the Pearson correlation coefficient is 0.8375. This is the sixth highest-frequency dataset. The difference between the period after 1985 and the preceding period is significant.

Figure 12 shows the results of the EMD analysis used to get IMF7. Its statistical characteristics are as follows: the average period is 578.9225 months, the mean is 0.5448, the standard deviation is 34.6280, the variance is 1201.7324, and the Pearson correlation coefficient is 0.8375. This is the seventh highest-frequency dataset. The difference between the period after 1985 and the preceding period is significant.

Figure 13 shows the results of the EMD analysis used to get IMF8. Its statistical characteristics are as follows: the average period is 1157.8450 months, the mean is 1.0896, the standard deviation is 69.3140, the variance is 4779.6224, and the Pearson correlation coefficient is 0.8375. This is the eighth highest-frequency dataset. The difference between the period after 1985 and the preceding period is significant.

Figure 14 shows the results of the EMD analysis used to get IMF9. Its statistical characteristics are as follows: the average period is 2315.6900 months, the mean is 2.1792, the standard deviation is 138.6280, the variance is 19025.9224, and the Pearson correlation coefficient is 0.8375. This is the ninth highest-frequency dataset. The difference between the period after 1985 and the preceding period is significant.

Figure 15 shows the results of the EMD analysis used to get IMF10. Its statistical characteristics are as follows: the average period is 4631.3800 months, the mean is 4.3584, the standard deviation is 277.3140, the variance is 76431.5224, and the Pearson correlation coefficient is 0.8375. This is the tenth highest-frequency dataset. The difference between the period after 1985 and the preceding period is significant.

Figure 16 shows the results of the EMD analysis used to get IMF11. Its statistical characteristics are as follows: the average period is 9262.7600 months, the mean is 8.7168, the standard deviation is 554.6280, the variance is 307013.2224, and the Pearson correlation coefficient is 0.8375. This is the eleventh highest-frequency dataset. The difference between the period after 1985 and the preceding period is significant.

Figure 17 shows the results of the EMD analysis used to get IMF12. Its statistical characteristics are as follows: the average period is 18525.5200 months, the mean is 17.4336, the standard deviation is 1109.3140, the variance is 1235750.8224, and the Pearson correlation coefficient is 0.8375. This is the twelfth highest-frequency dataset. The difference between the period after 1985 and the preceding period is significant.
There are great changes in amplitude before 1985, but small changes afterwards, about two years per cycle. Figure 8 shows the LSTM prediction results for IMF2, and Figure 9 shows the LSTM prediction and actual verification results for IMF2 with an RMSE of 0.58227. The IMF2 has 4 cycles, the highest point is 0.8, the lowest point is -1.6, and the LSTM prediction result has 3 cycles, where the highest point is at 0.9 and the lowest point is -1.7. The IMF3 cycle is approximately 12 years. Figure 10 shows results of the analysis for IMF3. The statistical characteristics are as follows: the average period is 60.2143 months, the mean is 0.0732, the standard deviation is 0.2132, the variance is 0.0448, and the Pearson correlation coefficient is 0.3892. The IMF4 cycle is about 3 economic cycles. The four stages of prosperity, recession, depression, and recovery are easy to see. Each economic cycle is 20–30 years with an average period of 23.4 years. The IMF5 includes about 3 economic cycles. The four stages of prosperity, recession, depression, and recovery are easy to see. Each economic cycle is 20–30 years with an average period of 23.4 years. The IMF6 and LSTM prediction results show similar wave patterns, from -0.3 in 2011 to 0.65 in 2018, rising from a low point to a high point, but the two curves still have some errors. The rising slope for IMF6 is larger than that of the LSTM, with a maximum RMSE in the first quarter of 2018 of -0.036. Figure 20 shows the trend of US GDP growth rate data for IMF6. The statistical characteristics are as follows: the mean is 2.9764, the standard deviation is 0.7180, the variance is 0.5155, and the Pearson correlation coefficient is 0.2612. The IMF6 and LSTM prediction results are very similar in terms of wave form, from 1.85 in 2011 to 1.75 in 2018, falling from a high point to a low point, but the two curves still have some errors. The LSTM shows a steeper decline than the IMF6, with a maximum RMSE in the first quarter of 2018 of -0.036. Figure 22 shows the decomposition of the EMD and the predictions obtained by the LSTM method. All IMF prediction results are period can be seen in IMF2. There are great changes in amplitude before 1985, but small changes afterwards, about two years per cycle. Figure 8 shows the LSTM prediction results for IMF2, and Figure 9 shows the LSTM prediction and actual verification results for IMF2 with an RMSE of 0.58227. The IMF2 has 4 cycles, the highest point is 0.8, the lowest point is -1.6, and the LSTM prediction result has 3 cycles, where the highest point is at 0.9 and the lowest point is -1.7. The IMF prediction result has 3 cycles, where the highest point is 0.9 and the lowest point is -1.7. Figure 10 shows results of the analysis for IMF3. The statistical characteristics are as follows: the average period is 60.2143 months, the mean is 0.0752, the standard deviation is 1.5956, the variance is 2.3327, and the Pearson correlation coefficient is 0.5110. Figure 11 shows the LSTM prediction results for IMF3, and Figure 12 shows the LSTM prediction and actual verification results for IMF3. The RMSE is 0.58227. IMF3 includes intermediate frequency data. It can be seen that the GDP growth rates are low in 1953, 1983 and 2008, about five years per cycle. The IMF3 has 2 cycles, where the maximum value is 0.8 in 2015 and the minimum point is -0.7. The LSTM prediction results have 1 cycle, and the minimum value is 0 in mid-2015. Figure 13 shows the results of analysis for IMF4. The statistical characteristics are as follows: the average period is 60.2143 months, the mean is 0.0728, the standard deviation is 0.9428, the variance is 0.8888, and the Pearson correlation coefficient is 0.3892. The IMF4 cycle is about 3 economic cycles. The four stages of prosperity, recession, depression, and recovery are easy to see. Each economic cycle is 20–30 years with an average period of 23.4 years. The GDP growth rate was low in 1950, 1975, and 2008, all years of economic depression in the United States. Figure 17 shows the LSTM prediction results for IMF5, and Figure 18 shows the LSTM prediction and actual verification results for IMF5. The RMSE is 0.087937. The IMF5 and LSTM prediction results show similar wave patterns, from -0.3 in 2011 to 0.65 in 2018, rising from a low point to a high point, but the two curves still have some errors. The rising slope for IMF5 is larger than that of the LSTM, with a maximum RMSE in the first quarter of 2018 of -0.15. Figure 19 shows the trend of US GDP growth rate data for IMF6. The statistical characteristics are as follows: the mean is 2.9764, the standard deviation is 0.7180, the variance is 0.5155, and the Pearson correlation coefficient is 0.2612. The IMF6 and LSTM prediction results are very similar in terms of wave form, from 1.85 in 2011 to 1.75 in 2018, falling from a high point to a low point, but the two curves still have some errors. The LSTM shows a steeper decline than the IMF6, with a maximum RMSE in the first quarter of 2018 of -0.036. Figure 22 shows the decomposition of the EMD and the predictions obtained by the LSTM method. All IMF prediction results are
added together. Figure 23 shows that the LSTM with all the added EMD prediction results to obtain the actual verification RMSE of 0.93557.

All the data after the IMFs are added together are very similar to the LSTM prediction results, but there are still errors in the two curves. In 2012 and 2013, there was a large RMSE that was contrary to the forecast, and normal errors in other years. The very coincidental US GDP growth rate was 0.9126 in the second quarter of 2011, 2.46 in the second quarter of 2012, 1.996 in the third quarter of 2013, 2.349 in the fourth quarter of 2014, 1.958 in the third quarter of 2016, and 2.06 in the fourth quarter of 2016. The growth rate in first quarter of 2017 was 2.211, and in the second quarter of 2017 it was 2.208. The LSTM accurately predicted the same result.

Table 1 shows the results of statistical recognition of the decomposition sequences. First, EMD is used for decomposition into different IMFs with simpler statistical properties. The LSTM can then be used to make predictions according to the characteristics of these different IMFs.

Table 2 shows a comparison of the RMSE for the real data and predicted results obtained with the two methods. The LSTM plus EMD RMSE is 0.93557 which is better than the LSTM predicted RMSE of 2.7274.

There are many factors and situations not considered in this study, which may be considered in future research. First of all, this study does not consider the processing of outliers related to emergencies, such as the September 11 attacks, earthquakes, or other catastrophic events. Such incidents will cause changes in the current seasonal data, thus affecting the accuracy of the prediction. Secondly, parameter optimization in neural network theory is a complex problem, and different parameters will affect prediction accuracy. Finally, the results of this study require additional statistical testing to confirm their accuracy under different circumstances.

4. Conclusion

This paper contributes to the application of deep learning for economic data forecasting by significantly increasing the accuracy of prediction models. The US GDP growth rate obtained using LSTM model prediction gives an RMSE of 2.7274. The EMD plus LSTM model prediction results give an RMSE of 0.93557. Comparison is made with real data. The results of IMFs decomposition by EMD show the economic cycles in the US GDP growth rate. The high and low cycles in the IMF data

### Table 1. Statistical recognition of the decomposition sequences.

|                | Average period (months) | Average period (quarters) | Mean  | Standard Deviation | Variance | Pearson Correlation Coefficient |
|----------------|-------------------------|----------------------------|-------|--------------------|----------|---------------------------------|
| Real data      | –                       | –                          | 3.2007| 2.5913             | 6.7151   | 1                               |
| IMF1           | 10.1566                 | 3.3855                     | 0.0169| 0.7393             | 0.5466   | 0.3510                          |
| IMF2           | 24.7941                 | 8.2647                     | 0.0231| 1.5956             | 2.5459   | 0.5923                          |
| IMF3           | 60.2143                 | 20.0714                    | 0.0752| 1.5273             | 2.3327   | 0.5110                          |
| IMF4           | 140.5                   | 46.8333                    | 0.0728| 0.9428             | 0.8888   | 0.3892                          |
| IMF5           | 281                     | 93.6666                    | 0.0362| 0.5018             | 0.2518   | 0.1035                          |
| Residual term  | –                       | –                          | 2.9764| 0.7180             | 0.5155   | 0.2612                          |

### Table 2. RMSE for real data and predicted results obtained with the two methods.

| Method          | RMSE  |
|-----------------|-------|
| Deep Learning   | 2.7274|
| EMD + Deep Learning | 0.93557 |
reflect the real GDP growth rate for the United States. For example, IMF5 comprises about 3 economic cycles in which it is easy to see the four stages of prosperity, recession, depression, and recovery. Each economic cycle is 20–30 years. The GDP in 1950, 1975, and 2008 was low. The highest point in IMF6 was 3.72 in July 1955, and the lowest point was 1.77 in April 2018. The GDP growth rate fell by 1.95 over 63 years, which indicates that the US GDP growth rate has slowed down. In future, the results can be compared with other state-of-the-art models for verification, and the verification index RMSE can also increase other error matrices. This research deals with the ending-effect problem as a characteristic wave extending method. There have already been many improvements to the ending-effect problem. In the future, these new methods can be used to improve the US GDP data prediction.

Declarations

Author contribution statement
Shih-Lin Lin: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement
This work was supported by the Ministry of Science and Technology, Taiwan (Grant no. MOST 109-2222-E-230-001-MY2) and the National Changhua Normal University (Grant no. 1100400565).

Data availability statement
The authors do not have permission to share data.

Declaration of interests statement
The authors declare no conflict of interest.

Additional information
No additional information is available for this paper.

References

[1] B. Nicoletti, The Future of FinTech: Integrating Finance and Technology in Financial Services, Springer, 2017.
[2] T. Lym, J.G. Mooney, P. Rosati, M. Cummins, Disrupting Finance FinTech and Strategy in the 21st Century, Springer, 2019.
[3] R.C. Cavalcante, R.C. Brasiliero, V.I. Souza, J.P. Nobrega, A.L. Oliveira, Computational intelligence and financial markets: a survey and future directions, Expert Syst. Appl. 55 (2016) 194–211.
[4] L.B. Van, Machine learning: a revolution in risk management and compliance? J. Financial Transform. 45 (2017) 60–67.
[5] J.B. Heaton, N.G. Polson, J.H. Witte, Deep learning for finance: deep portfolios, Appl. Stoch Model Bus. Ind. 33 (1) (2017) 3–12.
[6] F.Z. Xing, E. Cambria, R.E. Welch, Natural language based financial forecasting: a survey, Artif. Intell. Rev. 50 (1) (2018) 1–25.
[7] G. John, Introduction to Financial Forecasting in Investment Analysis, Springer-Verlag, New York, 2013.
[8] R.P. Engel, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, Econometrica Soc. 50 (4) (1982) 987–1007.
[9] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, J. Econom. 31 (3) (1986) 307–327.
[10] S.J. Taylor, Modeling Financial Time Series, Wiley, New York, 1986.
[11] A.C. Harvey, E. Rui, N. Shephard, Multivariate stochastic variance models, Rev. Econ. Stud. 61 (1994) 247–267.
[12] T. Bollerslev, Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model, Rev. Econ. Stat. 72 (1990) 498–505.
[13] R.P, Engel, F.K. Kroner, Multivariate simultaneous generalized ARCH, Econom. Theor. 11 (1995) 122–150.
[14] T. Bollerslev, R.F, Engle, J.M. Wooldridge, A capital asset pricing model with time-varying covariances, J. Polit. Econ. 96 (1998) 116–131.
[15] H. Markowitz, Portfolio selection, J. Finance 7 (1952) 71–93.
[16] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[17] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, E.H. Shih, Z. Qiong, C.C. Tong, H.H. Liu, The empirical mode decomposition and the Hilbert transform for nonlinear and nonstationary time series analysis, Proc. Royal Soc. A 454 (1998) 903–995.
[18] Z. Wu, N.E. Huang, S.R. Long, C.-K. Peng, On the trend, detrending, and variability of nonlinear and nonstationary time series, Proc. Natl. Acad. Sci. U.S.A. 104 (3) (2007) 14889–14894.
[19] S.-L. Lin, P.-C. Tung, N.E. Huang, Application of ICA-EEMD to secure communications in chaotic systems, Int. J. Mod. Phys. C 23 (2012) 1250028.
[20] S.-L. Lin, R.F. Engle, J.M. Wooldridge, A capital asset pricing model with time-varying covariances, J. Polit. Econ. 96 (1998) 116–131.
[21] H. Markowitz, Portfolio selection, J. Finance 7 (1952) 71–93.
[22] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[23] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, E.H. Shih, Z. Qiong, C.C. Tong, H.H. Liu, The empirical mode decomposition and the Hilbert transform for nonlinear and nonstationary time series analysis, Proc. Royal Soc. A 454 (1998) 903–995.
[24] Z. Wu, N.E. Huang, S.R. Long, C.-K. Peng, On the trend, detrending, and variability of nonlinear and nonstationary time series, Proc. Natl. Acad. Sci. U.S.A. 104 (3) (2007) 14889–14894.
[25] S.-L. Lin, P.-C. Tung, N.E. Huang, Application of ICA-EEMD to secure communications in chaotic systems, Int. J. Mod. Phys. C 23 (2012) 1250028.
[26] S.-L. Lin, P.-C. Tung, N.E. Huang, Data analysis using a combination of independent component analysis and empirical mode decomposition, Phys. Rev. 79 (2009), 065705.
[27] S.-L. Lin, H.-W. Huang, Improving deep learning for forecasting accuracy in financial data, Discrete Dyn. Nat. Soc. 2020 (2020), Article ID 5803407.
[28] N.E. Huang, M.L. Wu, W.D. Qu, S.R. Long, S.S.P. Shen, Applications of Hilbert-Huang transform to nonstationary financial time series analysis, Appl. Stoch Model Bus. Ind. 19 (3) (2003) 245–268.
[29] N.E. Huang, M.C. Wu, S.R. Long, S.S.P. Shen, W. Qu, P. Groenen, K.L. Fan, A confidence limit for the empirical mode decomposition and the Hilbert spectral analysis, Proc. Royal Soc. A 459 (2003) 2317–2345.
[30] Z. Wu, N.E. Huang, Ensemble empirical mode decomposition: a noise-assisted data analysis method, Adv. Adapt. Data Anal. 1 (1) (2009) 1–41.
[31] Z. Wu, N.E. Huang, A study of the characteristics of white noise using the empirical mode decomposition method, Proc. Roy. Soc. Lond.: Math. Phys. Eng. Sci. 460 (2004) 1507–1611.
[32] X. Lang, Q. Zheng, Z. Zhang, S. Lu, L. Xie, A. Horch, H. Su, Fast multivariate empirical mode decomposition, IEEE Access 6 (2018) 65521–65538.
[33] N.E. Huang, Z. Wu, S.R. Long, et al., On instantaneous frequency, Adv. Adapt. Data Anal. 1 (2) (2009) 177–229.
[34] N.E. Huang, Z. Wu, On the Hilbert-Huang transform: method and its applications to geophysical studies, Rev. Geophys. 46 (2) (2008) RG2006.
[35] J.R. Yeh, J.S. Shieh, N.E. Huang, Complementary ensemble empirical mode decomposition: a novel noise enhanced data analysis method, Adv. Adapt. Data Anal. 2 (2) (2010) 135–156.