The Utilization of a Recurrent Neural Network to Predict the Next U.S. Economic Recession

Abstract

In December of 2007, the housing bubble, worth $8 trillion, popped. This was catastrophic for the real estate sector of the economy and eventually led to the Great Recession. By creating a machine learning algorithm to determine the precise association between the indicators and recessions, researchers can create an effective method for predicting future economic calamities. Among several indicators of recessions, one signal stands out in particular: the yield curve. The yield curve is typically derived from the difference between the 3-month bond and the 10-year bond. Occasionally the yield curve will experience an economic phenomenon referred to as yield curve inversion. For the past 60 years, this yield curve inversion has acted as a precursor to economic recessions. To predict the occurrence of a recession, a recurrent neural network (RNN) was used. The first phase of research consisted of a program, Algorithm A, that predicted a historical recession, the Great Recession, to determine the effectiveness of the RNN. The second phase was a program, Algorithm B, that predicted future yield curve rates using historical yield curve data. Finally, the last phase of research, Algorithm C, was the integration of both algorithms to produce a single program that would use Algorithm A to predict the occurrence of a future recession based on the yield curve rates produced by Algorithm B. The results of this research predicted a recession to occur in November of 2020 and last until April of 2022. These results concur with a survey conducted by the National Association of Business Economics in which 72% of economists responded that they believed a recession would occur before the end of 2021 and 38% of economists predict a recession to occur in 2020. This prediction of a recession may be due to several factors including the business cycle, fiscal policy, trade tensions, and asset markets.

Introduction

In December of 2007, the housing bubble, worth $8 trillion, popped. This was catastrophic for the real estate sector of the economy and eventually led to the downfall of the global economy. In addition to home mortgage foreclosures, employment dropped 6.1% with the U.S. Labor Market reporting a loss of 8.4 million jobs. This, and other economic gauges, indicated the severity of the economic downturn. In fact, this period, lasting until June of 2009, was considered the most dramatic economic decline since the Great Depression and was named the "Great Recession" (Economic Policy Institute, n.d.). The United States economy still felt the effects of the Great Recession even six years following its beginning (Shierholz, 2014). Although the Great Recession was a devastating event for many, it provides economists, politicians, and statisticians information about the causes and effects of the event. In fact, researchers can learn from the all of the historic economic recessions to determine the causes and indicators of a recession. Several causes that have already been identified by economists include the unemployment rate, GDP output gap, stock market, housing market, manufacturing index, and several others (Ahmed, 2019). However, one signal of an upcoming recession stands out in particular: the yield curve.

The yield curve originates from the United States treasury-note prices. These are bonds issued by the federal government with varying maturity dates ranging from 1 month to 30 years. The yield curve is typically derived from the difference between the 3-month bond and the 10-year bond. As a result, economists can determine the disparity between short-term and long-term outlooks of the economy. In addition, researchers can collect information pertaining to federal interest rates and inflation rates. These, in turn, can affect real fiscal policy (Estrella & Mishkin, 1996). Naturally, the long-term bonds such as the 10-year or 30-year bonds outperform the short-term bonds such as the 1-month or 3-month bonds. This is intuitive as long-term bonds customarily offer high interest rates to entice potential investors.

However, occasionally the yield curve will experience an economic phenomenon referred to as yield curve inversion. This occurs when investors anticipate an oncoming economic calamity such as the lowering of interest rates or slower economic growth in general. Consequently, investors avoid long-term investments that can be risky and turn to short-term options. This results in short-term bonds outperforming the long-term bonds. As the name of this event implies, the yield curve inverts and the difference between the 3-month bond and the 10-year bond becomes negative. For the past 60 years, this yield curve inversion has acted as a precursor to economic recessions. In fact, it is considered one of the most reliable indicators of a recession (Phillips, 2019). Determining the exact start date and length of a recession, however, is not as simple.

Nevertheless, with machine learning, researchers are able to discover abnormalities, trends, and patterns in large datasets that humans cannot. These discoveries are ultimately used to make predictions that can have a surprising degree of accuracy. One particular machine learning algorithm is a recurrent neural network (RNN) which is intended for time-series data such as stock prices. This allows researchers to develop models that can recognize trends that occur over time. Among the many layers used to construct an RNN, the long short-term memory (LSTM) layer has gained increasing popularity with its application to several RNN algorithms such as natural language processing. The LSTM layer allows the RNN to remember information for longer periods of time which can increase the accuracy of the model.

As shown in Appendix B, the internal structure of an LSTM layer can become extremely complex. Since it is an RNN, the current cell begins by receiving information from the previous cell. This, combined with the new input data, are passed through a sigmoid function as part of the

“Forget Gate” (G_f). This allows the model to decide which pieces of information are necessary to determine the output and state of the cell. The sigmoid function is used to simplify the data by compressing the input to values between 0 and 1. Once this data is passed through the “Forget Gate” it is multiplied to the cell state vector. Next, the data is passed through another sigmoid function, as part of the “Update Gate” (G_u), where it used to determine the state of the cell. This, combined with a hyperbolic tangent function which is also used to simplify the data, are added to the cell state vector. The hyperbolic tangent function differs from sigmoid function in that it compresses the data into values between -1 and 1. Finally, the cell determines which pieces of information will passed to the next cell through the “Output Gate” (G_o). This is accomplished by using another set of sigmoid and hyperbolic tangent functions (Hoffman, 2018).

The data presented by the yield curve and its relationship to impending recessions makes it an exceptional candidate for research with RNNs. By creating a machine learning algorithm to determine the precise association between the yield curve and recessions, researchers can create an effective method for predicting future economic calamities. This research project was divided into three distinct phases. The first phase consisted of a program, Algorithm A, that predicted a historical recession, the Great Recession, to determine the effectiveness of the RNN. The second phase was a program, Algorithm B, that predicted future yield curve rates. This was completed using historical yield curve data. Finally, the last phase of research, Algorithm C, was the integration of both algorithms to produce a single program that would use Algorithm A to predict the occurrence of a future recession based on the yield curve rates produced by Algorithm B.

Methods and Materials

This research was conducted on a HP Pavilion Laptop with an Intel i3 core processing chip. The programming language utilized for this research project was Python 3.7.3 and the integrated development environment (IDE) used was PyCharm. Before beginning the program, several packages such as Numpy and Pandas were installed for the data handling and manipulation. For the machine learning algorithm packages, Keras and TensorFlow were installed. Finally, Anaconda, a package management system, was installed to the laptop. The yield curve data used for this research was downloaded in a comma separated value (csv) format from the Treasury Yield Curve Rates by the U.S. Treasury. This data was published on Quandl, an open-data site for economic, financial, and alternative data. The data had monthly intervals and ranged from January of 1990 to October of 2019. However, this dataset did not include information regarding the occurrence of a recession. In Microsoft Excel, appropriate values were added in a separate column to denote the occurrence of a recession for each data point. A "0" or "1" represented "no recession" or "recession", respectively.

Following the data preparation, phase one of the research and the creation of Algorithm A began. This commenced with reading in the file to the program. Then, the yield curve was calculated by subtracting the 3-month rates from the 10-year rates. Then, the recession values were stored in a separate variable. Next, the values were stacked and separated into training data, approximately 50%, and testing data, approximately 50% to ensure that the entirety of the Great Recession would be placed in the testing dataset. Then the data was scaled to minimize the amount of loss within the machine learning training process. Then the data was split into x-values and y-values. This was done by appending the first 50 yield curve rates to an array of x-values and 1 recession value from the 50^{th} index to an array of y-values. This frame was shifted and repeated resulting in 129 frames of 50 datapoints. Once both of the training data and the testing data had been split into x-values and y-values, the data was reshaped to be passed into the RNN.

As shown in Appendix C, the architecture for the neural network was as follows: a LSTM layer, dropout layer, LSTM layer, dropout layer, LSTM layer, dropout layer, LSTM layer, dropout layer, dense layer. Each LSTM layer had 50 units or neurons and the dense layer contained one neuron. Dropout layers were utilized to drop some data generated by the model. This prevents the model from becoming overloaded with computation. After the model had been constructed, it was compiled using "mean squared error" for the loss function and "adam" as the optimizer. Then, the model was trained with the training data for 50 epochs, or iterations, and a batch size of 43. After the model trained, it was tested with the testing data. The predictions were rounded and stored in an array. The program then found the start index and end index of all consecutive 1's. These indices, along with the original yield curve data, were used to graph the results of the model. Using matplotlib.pyplot, the data was graphed.

The next phase was to create Algorithm B. This too began with reading the data as a Pandas data frame and calculating the yield curve by subtracting the 3-month rates from the 10-year rates. Then the data was split into training data, approximately 85%, and testing data, approximately 15%. This was done to ensure that the testing dataset consisted of exactly 95 datapoints, the selected model input size. Next, the data was split into x-values and y-values. This was similar to the process in Algorithm A with one notable exception: Algorithm B was designed to make a multi-step prediction. So, 95 yield curve values were appended to an array of x-values and the next 50, instead of 1, yield curve values were appended to an array of y-values. The frame was shifted, and process was repeated. After reshaping the data, it was passed into the RNN. The architecture for the neural network was the same as Algorithm A except that it had 95 neurons per LSTM layer and 50 neurons in the dense layer. After the model had been constructed, it was compiled using "mean squared error" for the loss function and "adam" as the optimizer. Then, the model was trained with the training data for 50 epochs and a batch size of 17. Next, the model was used to predict yield curve data for the next 50 months using 95 yield curve datapoints in the testing data. The results were stored and plotted with the original yield curve data to represent the new yield curve data generated by the RNN (Appendix G).

The third phase of this research was the implementation of Algorithm C. In one file, Algorithm A and B were utilized to attain this goal. After the future data had been generated by the Algorithm B, it was concatenated to the historic yield curve data to create the training data for Algorithm A. To ensure that the yield curve column and the recession column had the same length, zeros were added to the recession value column. Then Algorithm A was implemented to make a recession prediction based on this new set of data. The training data was approximately 80% and the testing data was about 20% of the concatenated yield curve data. This would allow the model to have more training data, including data for the Great Recession, and, in turn, have greater accuracy.

Results

As shown in Figure 1, the results of this research predicted a recession to occur in November of 2020 and last until April of 2022. The future data generated by Algorithm B showed a significant increase in yield curve price after subzero rates, further indicating the possibility of a recession (Appendix E). Algorithm A predicted the Great Recession to begin in February of 2007 and end in February of 2008, whereas the actual dates of the recession were December of 2007 to June of 2009 (Appendix D). With the addition of more training data however, the accuracy of Algorithm C is believed to be higher than that of the Algorithm A.

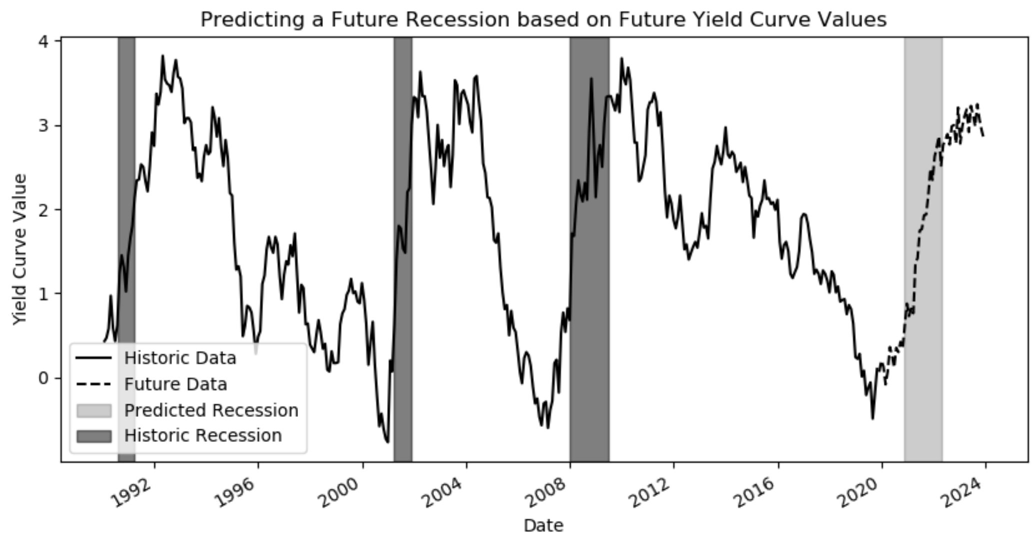


Figure 1. Predicting a Future Recession based on Future Yield Curve Rates (Algorithm C)

Conclusion

In this research, an RNN was developed to utilize yield curve rates to predict a future recession. As the results indicated, the program predicted a recession to occur from November of 2020 to April of 2022. The implementation of the Algorithm A did serve as a proof-of-concept to demonstrate that the RNN was able to identify a historic recession. This indicates that the RNN performed considerably well. In addition, Algorithm B produced reasonable values when comparing to past yield curve rates.

In a survey conducted by the National Association of Business Economics, 72% of economists responded that they believed a recession would occur before the end of 2021 and 38% of economists predict a recession to occur in 2020 (Heeb, 2019). These statistics further prove the validity of the results generated by the RNN. Overall, the significant difference between the number of economists who predict a recession to occur before 2021 and those who predict one to occur before 2020 illustrates that dates produced by the machine learning model may be early. In fact, the historical prediction from the Algorithm A was slightly earlier than the actual time period of the Great Recession.

This issue, and others, with the program may have resulted from a number of problems within the code. A common complication with machine learning algorithms is the lack of data. Although the data set for this program had nearly 30 years of data and over 350 individual data points, it may have been insufficient for the program to accurately predict a recession. Another potential issue was overfitting. This occurs when a model becomes too familiar with the training data and simply memorizes a specific sample. This may have occurred in the RNN training process.

However, the model was able to predict the occurrence of a recession. This recession may be due in part by a number of events. For now, United States is experiencing one of the longest economic expansions in its history (nearly ten years). This is also why some economists believe there will be an upcoming recession; the cyclical nature of the economy guarantees bull markets and bear markets. Other reasons for a recession include the continuing trade war with China (Irwin, How the Recession of 2020 Could Happen, 2019). Also, the Federal Reserve may play a role in determining a recession. In the past, they have overestimated growth and hiked interest rates to combat hyperinflation. However, this overestimation leads to a rapid reduction in economic growth and, in turn, a recession (Bivens, 2019). Finally, an economic recession may arise from the popping of an asset bubble. In both 2001 and 2007, the dot-com bubble and the housing bubble popped, respectively (Irwin, What Will Cause the Next Recession? A Look at the 3 Most Likely Possibilities, 2018). All of these factors, and more, may have likely contributed to the yield curve inversion, and ultimately, a future recession.

References

Peer-Reviewed

Bivens, J. (2019, April 18). *What should we know about the next recession?* Retrieved from Economic Policy Institute: https://www.epi.org/publication/next-recession-bivens/

Economic Policy Institute. (n.d.). *The Great Recession*. Retrieved from The State of Working America: http://stateofworkingamerica.org/great-recession/

Estrella, A., & Mishkin, F. S. (1996, June). The Yield Curve as a Predictor of U.S. Recessions. *Current Issues in Economics and Finance, 2*(7), pp. 1-2.

Hoffman, G. (2018, January 11). *Introduction to LSTMs with TensorFlow*. Retrieved from O'Reilly: https://www.oreilly.com/ideas/introduction-to-lstms-with-tensorflow

Shierholz, H. (2014, January 9). *Six Years from Its Beginning, the Great Recession's Shadow Looms Over the Labor Market*. Retrieved from Economic Policy Institute: https://www.epi.org/publication/years-beginning-great-recessions-shadow/

Non-Peer-Reviewed

Ahmed, S. I. (2019, August 13). *Predicting the next U.S. recession*. Retrieved from Reuters: https://www.reuters.com/article/us-usa-economy-watchlist-graphic/predicting-the-next-u-s-recession-idUSKCN1V31JE

Heeb, G. (2019, August 19). *More than 70% of economists think a US recession will strike by the end of 2021*. Retrieved from Markets Insider: https://markets.businessinsider.com/news/stocks/next-recession-trump-survey-shows-most-economists-predict-downturn-2021-2019-8-1028456445

Irwin, N. (2018, August 2). *What Will Cause the Next Recession? A Look at the 3 Most Likely Possibilities*. Retrieved from The New York Times: https://www.nytimes.com/2018/08/02/upshot/next-recession-three-most-likely-causes.html

Irwin, N. (2019, August 17). *How the Recession of 2020 Could Happen*. Retrieved from The New York Times: https://www.nytimes.com/2019/08/17/upshot/how-the-recession-of-2020-could-happen.html

Phillips, M. (2019, November 8). *A Recession Warning Reverses, but the Damage May Be Done*. Retrieved from The New York Times: https://www.nytimes.com/2019/11/08/business/yield-curve-recession-indicator.html

Data Source

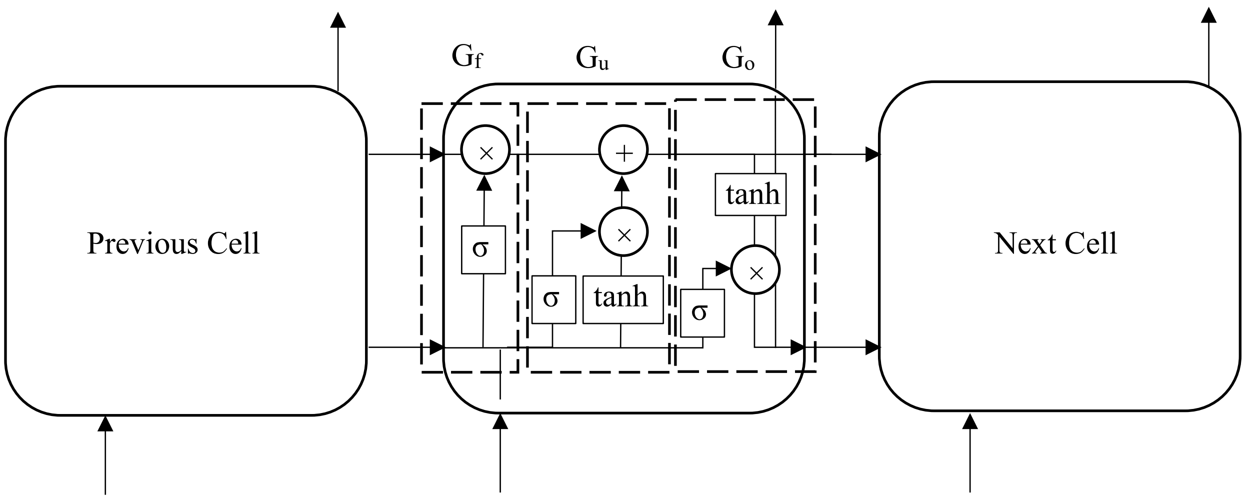
US Treasury. (2019). Treasury Yield Curve Rates. Quandl. Retrieved from https://www.quandl.com/data/USTREASURY/YIELD-Treasury-Yield-Curve-Rates

Appendix

Appendix A:

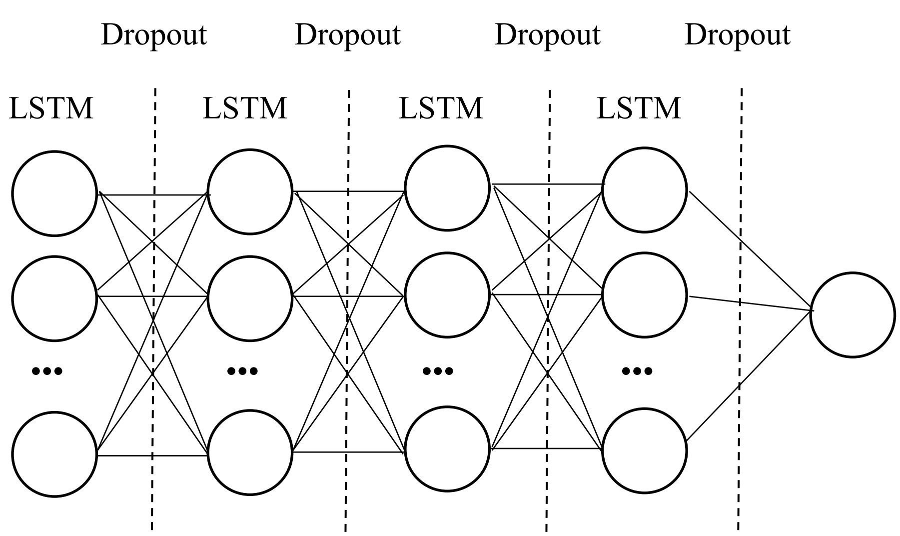
Program Code: https://github.com/josh1227/RecessionPrediction

Appendix B:



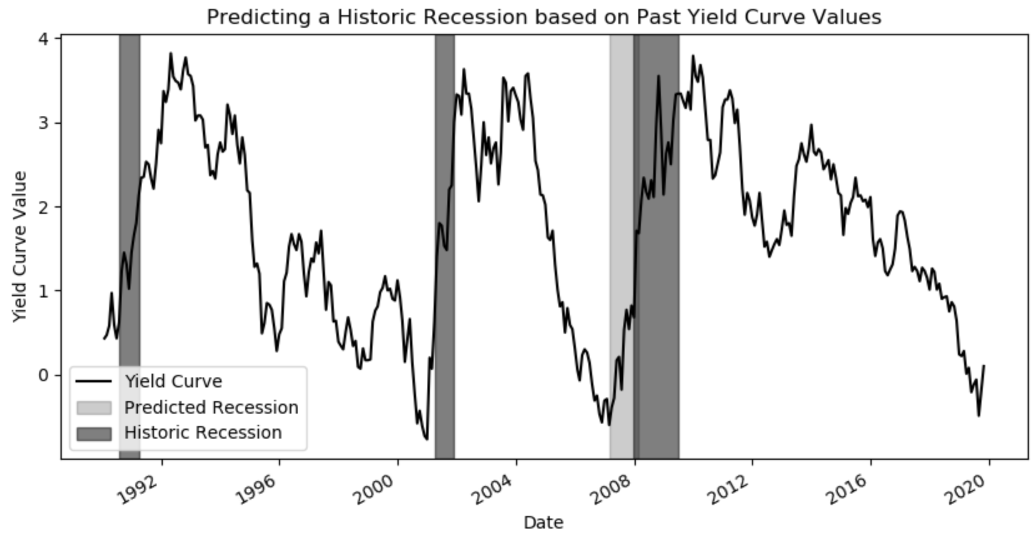
Appendix B. Recurrent Neural Network/Long Short-Term Memory Layer Structure

Appendix C:



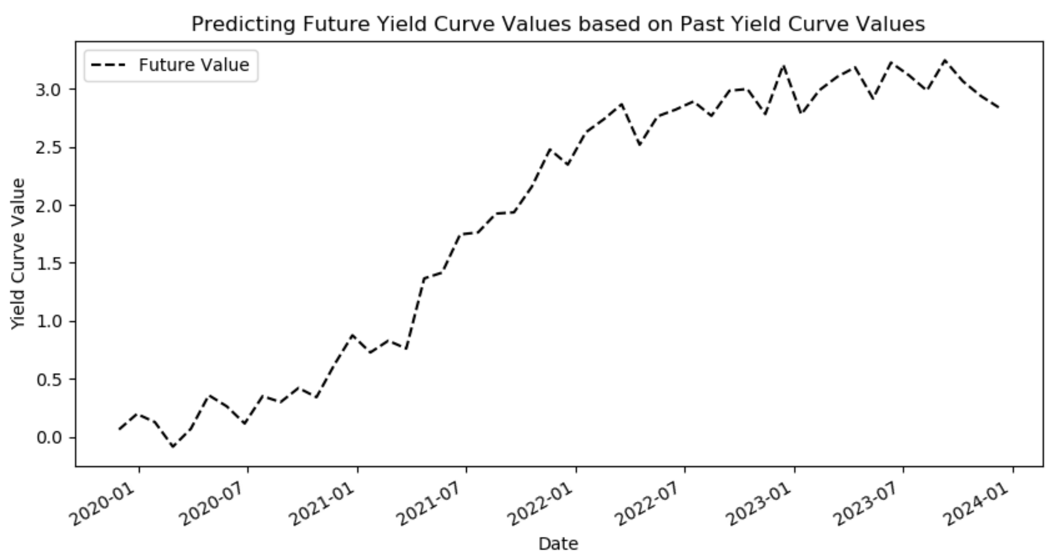
Appendix C. Recurrent Neural Network Architecture

Appendix D:



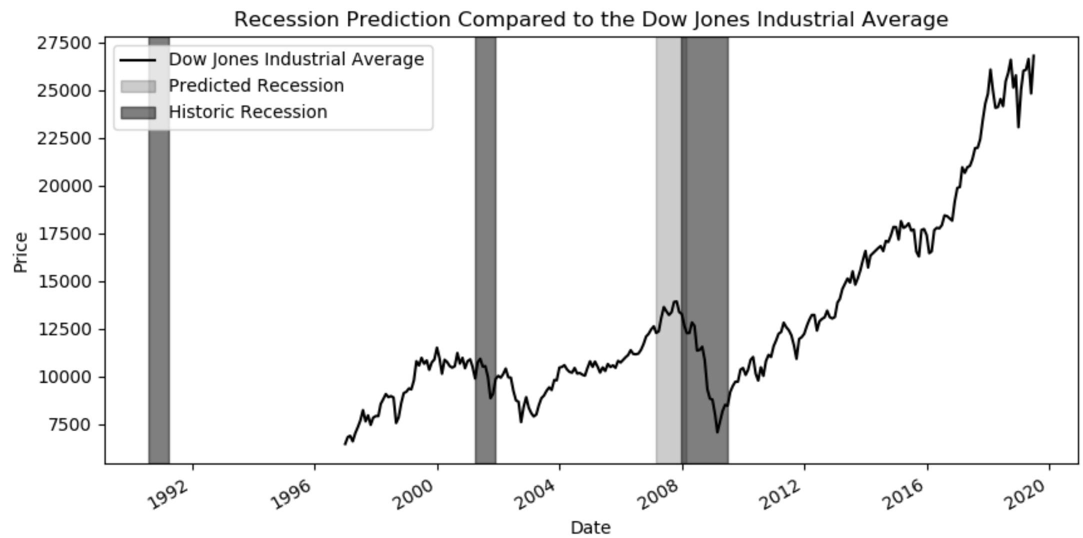
Appendix D. Predicting Historic Recession based on Past Yield Curve Rates (Algorithm A)

Appendix E:



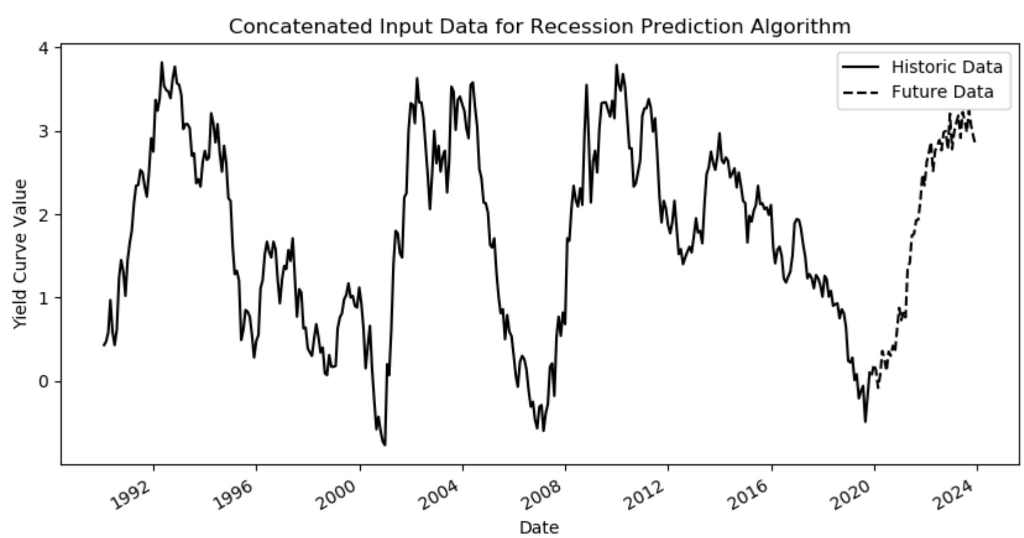
Appendix E. Predicting Future Yield Curve Values based on Past Yield Curve Rates (Algorithm B)

Appendix F:



Appendix F. Historic Recession Prediction Compared to Dow Jones Industrial Average

Appendix G:



Appendix G. Concatenated Input Data for Recession Prediction Algorithm