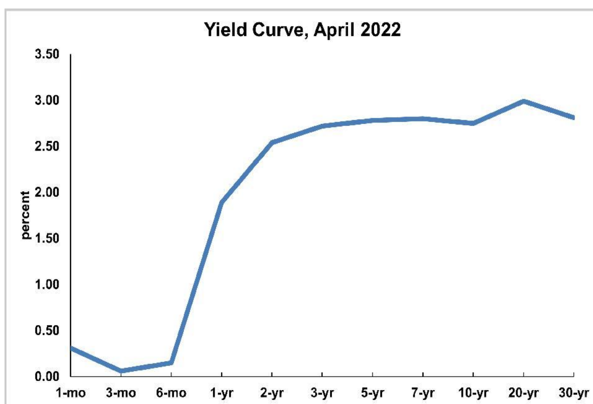
Is the Yield Curve a Leading Indicator of a Recession in the US?

Himabindu Thota, Khakali Olenja, Mima Mirkovic

Background

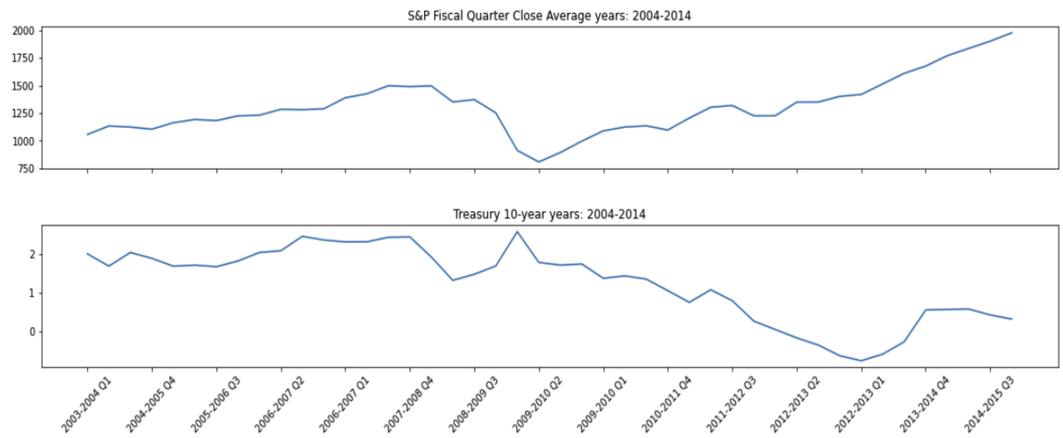
Our team would like to analyze both economic and financial data from 1990 through 2021 to predict the likelihood of a recession in 2022. We will analyze the S&P 500 Index and the US Treasury Daily Yield Curve Rates over that period of time to identify the indicators weighted most heavily for a recession and apply them to the conditions of fiscal trends in 2022. If successful, we will be able to create a probabilistic model that will enable us to predict the conditions that would indicate when a recession is most likely. Understanding if we are heading into a recession or already in one will have broad implications for individuals living in the US and the world at large.

What is the Yield Curve?
The Yield Curve is an economic indicator that depicts the relationship between short-term and long-term interest rates and the duration for US Treasury Notes, aka US Debt (Conerly, *Forbes*).



The yield curve is central to the transmission of monetary policy and provides information on future interest rates, inflation, and economic growth that aids banks, investors, and economists alike (*Reserve Bank of Australia*). Economists have long debated how useful monitoring the yield curve is at predicting a recession.

The three key types of yield curves include: *Normal, Inverted*, and *Flat*. Each has implications for the economy at large (Birken, *Forbes*). A **normal** yield curve indicates economic prosperity, where yields are increasing in maturity. An **inverted** yield curve implies a recession, where bond prices and yields gradually decrease. A **flat** yield curve implies flat economic growth, where short and long-term maturities boast similar yields.



What is the S&P 500 Index?

The S&P 500 Index is a weighted index of 500 leading publicly traded companies, and serves as an indicator of overall stock market performance. Faced with the threat of recession, the Federal Reserve increases interest rates to combat inflation, which in turn affects the behavior of investors and consumers alike (*Forbes*) Therefore, while recessions are viewed as economic downturns, the behavior of the stock market nonetheless has implications for the overall health of both the stock market and the economy.

Problem Statement

Since 1960, an inverted yield curve has preceded each recession. During a recession, short-term rates fall due to the Fed easing up on monetary policy and weakening credit, in turn negatively influencing investor behavior and thus resulting in an inverted yield curve (Stojanovic, *Federal Reserve Bank of St.Louis*).

As of today, inflation and increases in interest rates have consumers, investors, and economists alike pondering over the possibility of an economic downturn, with many claiming that we have already entered a recession. Given past and current economic trends, our goal is to back these claims by answering the question that everyone is trying to answer:

Can we analyze prior recessions to accurately predict the next one?

Objective

Our ultimate goal is to accurately predict an oncoming recession through machine learning. Since 1990, there have been four recessions through 2020. We will use these recessions as our baseline to predict for 2022:

- July 1990 - March 1991
- March 2001 - November 2001
- December 2007 - June 2009
- February 2020 - April 2020

Datasets

	S&P 500 Daily Data	U.S. Department of Treasury Daily Yield Curve Rates
Data	Historical data of S&P 500 Index from 1927 to 2021	Daily Treasury Par Yield Curve Rates from 1990 to 2021
Columns	Date, Open, High, Low, Close, Adj Close, Volume, % Gain/Loss, % Price Variation	Date, 1 Mo, 2 Mo, 3 Mo, 6 Mo, 1 Yr, 2 Yr, 5 Yr, 7 Yr, 10 Yr, 20 Yr, 30 Yr
Type	Daily; data collected every day except weekends and holidays	Daily; data collected every day except weekends and holidays
Entries	9 columns, 19685 non-null entries No missing values	13 columns, 8008 entries Missing values (total) : 12051 Nan values
2022 Data for Prediction	WSJ S&P 500 Index 2022 Prices Date, Open, High, Low, Close	Daily Treasury Par Yield Curve Rates for 2022 Same Variables

S&P 500 Daily Data (1927-12-30 to 2021-09-19) | Kaggle
Interest Rates Data CSV Archive | U.S. Department of the Treasury
SPX | S&P 500 Index Historical Prices - WSJ

Methodology

1. Data Collection
    - Compile data for analysis
    - Read-in data
    - Describe data
2. Exploratory Data Analysis
    - Conduct initial investigations on the data
    - Present bivariate and univariate representations of the data
    - Find and address missing values, clean data, prepare data for further analysis
3. Model Building
    - Split data into train, validation, and test sets
    - Select algorithm for model building
4. Model Training
    - Fit the model on the training data
    - Determine set of weights and biases with a low loss across all values
5. Model Evaluation

    - Tune and evaluate model parameters
    - Determine a working model that can be validated, tested, and deployed
6. Model Selection
    - Evaluate each model and draw comparisons
    - Select the appropriate model to make official predictions on
7. Predict
    - Make predictions on the test set using the selected model
    - Predict on previous years and on 2022
    - Plot predictions and visualize results
8. Interpret Results
    - Analyze final predictions and draw conclusions
    - Discuss constraints and limitations
    - Discuss model implementations and following studies

Approach

Independent Variables

U.S. Department of Treasury Daily Yield Curve Rates

- 3 Mo, 6 Mo: monthly yield curve rates
- 1 Yr, 2 Yr, 3 Yr, 5 Yr, 7 Yr, 10 Yr, 30 Yr: yearly yield curve rates
- *Difference*: numerical difference between the 10-year and 2-year rates

S&P 500 Daily Data

- High: Highest price during the trading day
- Low: Lowest Price during the trading day
- Close: Price at market closing
- Adjusted Close: Closing price after corporate actions are accounted for
- Volume: Number of shares traded during the day
- % Gain/Loss (Close): Percentage Change between 2 consecutive closing prices
- % Price Variation: Price fluctuation between High and Low during the day

Target Variable

We want our model to output a 0 or a 1, which classifies the results into 'No Recession' or 'Yes Recession' categories. We manually created a *'Recession'* variable that will serve as the target variable. Within this column, we manually assigned 0s and 1s to each day from 1990 to 2021 depending on whether or not that day reflected a recession. Our X variable is the data frame without the 'Recession' column and the y variable is the 'Recession' variable.

Splitting the Data

Features from both datasets will be combined for train and test sets. We want our model to learn from the majority of the data, or 80%. We will dedicate the remaining 20% to the test set. For the Neural Networks model, we will use a traditional stratified train/test split. For Logistic Regression, Random Forest and Ensemble, the data will be split manually by specific dates.

Prediction Approach

After conducting experiments and evaluating each model, we will select our candidate model. Using our chosen model, will first predict on the original dataset to ensure that the model is correctly classifying prior recessions. Then, we will add the 2022 data and create a new dataset to train and test our model on.

Algorithms for Model Building

Our project falls under the classification category. We want our model to output a 0 or 1 as well as classify the input data as a recession based on previous patterns. Therefore, we will consider algorithms that fall under the supervised learning approach, where we will train on labeled input data to build our predictive model. The algorithms we will explore are as follows: **Logistic Regression**, **Random Forest**, **Recurrent Neural Networks**, and **Ensemble**.

Success Criteria

Given the scope of our research and our insistence on building an effective model, we decided that an Accuracy Score greater than 80% is an acceptable baseline for both building and selecting our candidate model. We will consider the accuracy score as well as additional evaluation parameters for each respective model before deciding on our final model for predicting a recession.

Evaluation Parameters

Logistic Regression	Random Forest	Neural Networks	Ensemble
Accuracy Score Confusion Matrix	Accuracy Score Confusion Matrix Precision, Recall, F-1 Score, Support	Loss, Accuracy Score Confusion Matrix Precision, Recall, F-1 Score, Support	Accuracy Score Confusion Matrix

Exploratory Data Analysis

Data Cleaning

We first looked for any missing values. We found that '30 Yr' from the Daily Treasury Yield Curve Rates dataset had 995 missing values, while the other columns had only a few missing values collectively. We decided to impute these values using a *weekly average*, for which we had to write a separate algorithm to calculate these values. In cases where there is data missing for two or more contiguous weeks, we used the weekly average of the week occurring immediately before the weeks with no data. Rather than using the average of the entire column, we felt that computing the weekly average would be more specific and therefore accurate.

Date	3 Mo	6 Mo	1 Yr	2 Yr	3 Yr	5 Yr	7 Yr	10 Yr	30 Yr	Recession	diff	High	Low	Close	Adj Close	Volume	% Gain/Loss (Close)	% Price Variation
1990-01-02	7.83	7.89	7.81	7.87	7.90	7.87	7.98	7.94	8.00	0	0.07	359.69	351.98	359.69	359.69	162070000	1.7487	0.021
1990-01-03	7.89	7.94	7.85	7.94	7.96	7.92	8.04	7.99	8.04	0	0.05	360.59	357.89	358.76	358.76	192330000	-0.2592	0.008
1990-01-04	7.84	7.90	7.82	7.92	7.93	7.91	8.02	7.98	8.04	0	0.06	358.76	352.89	355.67	355.67	177000000	-0.8688	0.017
1990-01-05	7.79	7.85	7.79	7.90	7.94	7.92	8.03	7.99	8.06	0	0.09	355.67	351.35	352.20	352.20	158530000	-0.9852	0.012
1990-01-08	7.79	7.88	7.81	7.90	7.95	7.92	8.05	8.02	8.09	0	0.12	354.24	350.54	353.79	353.79	140110000	0.4494	0.010

Univariate and Bivariate Analysis

After cleaning the data and combining the datasets, we performed univariate and bivariate analyses on the datasets. The 2-year and 10-year rates had a high positive correlation (0.9338), and that the High, Low, Close, and Adjusted Close variables were positively correlated with the Volume. Since the Open variable was similar in both meaning and correlation to the Close and Adjusted Close variables, we decided to omit that variable from our final dataset. The histograms and accompanying scatter plots provided insight into the distribution of our data post processing. We found no significant irregularities with our data and the majority of our variables were normally distributed and ready for further analysis.

Experiments & Results

01 | Logistic Regression

Overview

Logistic Regression is useful for predicting the probability of a target variable. It is mainly used for binary classification purposes and is easier to implement, interpret, and efficient to train as opposed to other machine learning techniques.

Building the Model

We called the Logistic Regression algorithm and then fitted the model on the training data.

Results

Accuracy Score	`0.9566130160951715`
Confusion Matrix	`True Positives(TP) = 1367` `False Positives(FP) = 0` `True Negatives(TN) = 0` `False Negatives(FN) = 62`

02 | Random Forest

Overview

Random Forest is effective for prediction purposes and can handle large amounts of data. Opposed to the Decision Tree algorithm, Random Forest provides a higher degree of prediction accuracy. The Classifier algorithm uses bagging and feature randomness to generate a ‘forest’ whose predictions are more accurate than that of any individual tree.

Building the Model

We called the Random Forest Classifier algorithm using 200 estimators. We started at 1000 estimators and worked down until we found the optimal number of estimators.

Results

Accuracy Score	0.9545136459062281
Confusion Matrix	False Positives(FP) = 3 True Positives(TP) = 1364 False Negatives(FN) = 62 True Negatives(TN) = 0
Classification Report	precision recall f1-score support 0 0.96 1.00 0.98 1367 1 0.00 0.00 0.00 62 accuracy 0.95 1429 macro avg 0.48 0.50 0.49 1429 weighted avg 0.92 0.95 0.93 1429

03 | Recurrent Neural Network - LSTM Sequential Model

Overview

Neural networks are well equipped for predictive analysis due to their hidden layers, and allow us to interpret the data’s sequential characteristics to find patterns. Sequential data, such as stock prices, is best analyzed under a Long Short-Term Memory (LSTM) model, which retains memory from earlier inputs with each time-step. LSTM models are an improved version of a Recurrent Neural Network, as they have greater parameter flexibility, a higher degree of prediction accuracy, and can analyze multiple layers in parallel.

Splitting the Data

We used sklearn’s train/test split to allocate 80% of the data into a training set and the remaining 20% into a test and validation set. For this particular model, we needed to scale and reshape the sequential data in order to ensure that the network learned efficiently.

Building the Model

We used the traditional LSTM(128) for the first 3 layers, then used a fully connected Dense(32) layer with a sigmoid activation and a Dense(2) layer with a softmax activation in the final stages. Sigmoid produced a higher accuracy as opposed to a tanh activation. The final Dense(2) commonly uses a softmax layer to convert the scores to a normalized probability distribution. We used Dropout(0.2) to randomly exclude inputs at 20%, which returned a higher accuracy score as opposed to a higher Dropout rate. We compiled the model using the Adam Optimizer, which works well with large volumes of data. Additionally, we used Sparse Categorical Cross Entropy and the Accuracy metric, which work for binary classification purposes. We then fit and ran the model on ten epochs.

Results

Loss, Accuracy	`loss: 0.0759 - accuracy: 0.9622`
Confusion Matrix	False Positives(FP) = 17 True Positives(TP) = 1404 False Negatives(FN) = 43 True Negatives(TN) = 122
Classification Report	precision recall f1-score support 0 0.97 0.99 0.98 1421 1 0.88 0.74 0.80 165 accuracy 0.96 1586 macro avg 0.92 0.86 0.89 1586 weighted avg 0.96 0.96 0.96 1586

04 | Ensemble - LGBM Classifier

Overview

Gradient Boosting models are used for solving both regression and classification problems. They make a prediction model based on an ensemble of weak prediction models which are typically decision trees. We used the LightGBM ensemble model, which is a distributed and high-performing gradient boosting learning algorithm, as opposed to another gradient boosting algorithm such as XGBoost.

Building the Model

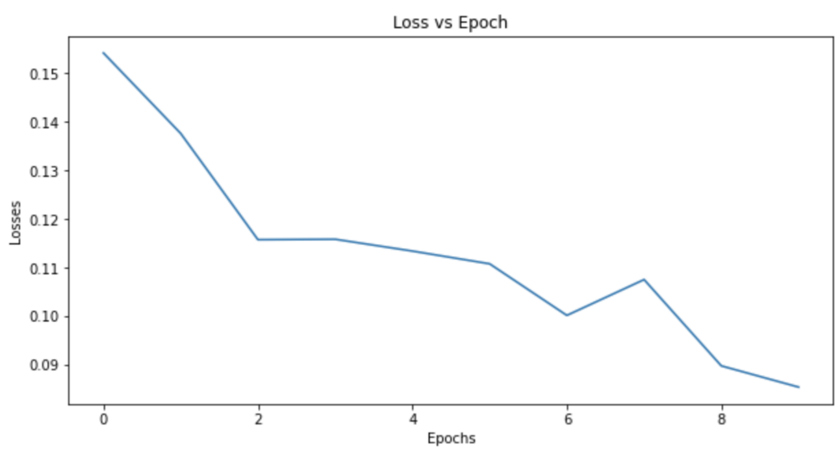
We called the lightgbm Classifier algorithm to train and then test our classification model.

Results

Accuracy Score	Test set score: 0.9566
Confusion Matrix	True Positives(TP) = 1367 False Positives(FP) = 0 True Negatives(TN) = 0 False Negatives(FN) = 62
Classification Report	precision recall f1-score support 0 0.96 1.00 0.98 1367 1 0.00 0.00 0.00 62 accuracy 0.96 1429 macro avg 0.48 0.50 0.49 1429 weighted avg 0.92 0.96 0.94 1429

Model Selection

After considering each model, we selected the RNN-LSTM Sequential Model as our final model. The model had a loss of 0.0759 and the highest accuracy score of 0.9622. The classification report had a high precision and recall score for both 0 and 1 (>0.74), and the confusion matrix returned 1404 True Positives, the highest amount of true positives for any model.



Out of the evaluation parameters, we believed that an LSTM model was valuable when it came to dealing with time-series data. An LSTM model can triage the important patterns and sift through large amounts of data, and it is well-equipped to classify, process, and predict.

Predictions

We predicted on the original dataset, without the 2022 data, and outputted the results in a data frame that included the predicted value and the actual value. The model was able to confirm that we had recessions in 1990, 2001, 2008, and 2020.

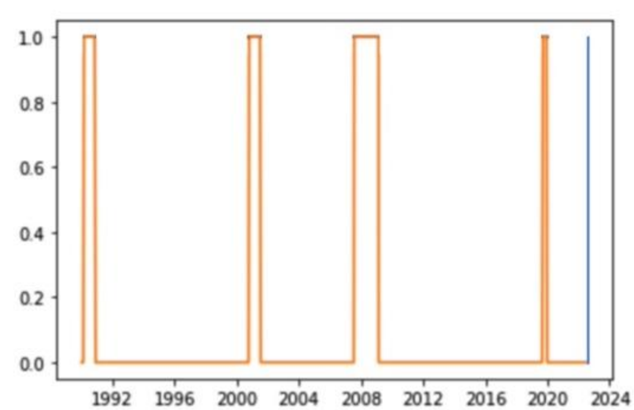
Date	Predicted	Actual
2020-02-03	0	1
2020-02-20	0	1
2020-02-24	0	1
2020-03-03	1	1
2020-03-10	1	1
2020-03-11	1	1
2020-03-20	1	1

Date	Predicted	Actual
1990-09-20	1	1
1990-09-21	1	1
1990-10-04	1	1
1990-10-09	1	1
1990-11-06	0	1
1990-11-08	1	1
1990-11-16	1	1

We added the 2022 treasury yield curve rates and the S&P 500 2022 prices, but the new S&P 500 data only had the "Open", "High", "Low", and "Close" variables. Rather than imputing any values, we decided to omit "Adjusted Close", "Volume", "% Gain/Loss", and "% Price Variation" from the new dataset in order to focus solely on the values that directly reflected the price of the stock. We also added the man-made "Recession" and "Difference" variables, filling the "Recession" column for the 2022 dates with all 0s to determine if the model could output 1s even after training on all 0s.

In order to predict ahead and potentially classify days in 2022 with 1s, we shifted the "Recession" data back one quarter and created a new variable called "one_q". This allowed the model to learn on the same parameters while a recession was potentially classified differently for each date. Using the same steps to scale and reshape the augmented dataset, we fit the model again and ended up with a 0.9831 validation accuracy after ten epochs.

Date	2022 Prediction	Actual
2022-05-02	1	0.0
2022-05-13	1	0.0
2022-05-16	1	0.0
2022-05-18	1	0.0
2022-06-08	1	0.0
2022-06-13	1	0.0
2022-06-14	1	0.0
2022-07-01	1	0.0
2022-07-12	1	0.0



Discussion

Ultimately, we found that the model classified dates in the previous quarter of 2022 with 1s, despite the 0s we used to fill the empty "Recession" values for 2022. In other words, the US economy appears to already be in a recession! Outside of our model, economists claim that a country is in a recession when gross domestic product (GDP) falls for two consecutive quarters back-to-back (Torabi, *CNET*). As it stands, GDP dropped by 1.6% in Q1 and 0.9% in Q2, information that points to a recession despite the US not officially declaring it (Torabi, *CNET*). With interest rates on the rise, an unstable stock market, and a declining GDP, we can turn to our model to confirm what the American population is already suspecting.

Constraints & Biases

In terms of time, LSTM models take longer to train and each layer requires a large amount of memory bandwidth to be computed. Additionally, LSTM models are easy to overfit, and even though this issue can be combated with the Dropout method, the model is still susceptible. In terms of scope, the question we are asking is both difficult and nearly impossible to answer. We were able to confirm what has already happened using predictive modeling, but predicting what's ahead is an entirely separate dilemma that even the most complex of models cannot solve. The limitation of our model is that it is applicable when we have the information readily available, not when we have to generate an unseen value.

With any machine learning model, it is difficult to replicate the human decision making process in an unbiased way. The new S&P 500 data was missing four variables that we used in the original dataset (Adjusted Close, Volume, % Gain/Loss, % Price Variation), and that omission could have introduced bias into our neural net. Additionally, the manual assignment of 0s and 1s as well as the computation of a weekly average could have introduced bias to the model. On the other hand, efforts to mitigate bias included: selection of a comprehensive dataset, thoroughly pre-processing the data, and selecting the most appropriate model for our data.

Standards

Python Version: Python 3.8.8

Packages & Modules
Pandas Numpy Time, datetime Matplotlib.pyplot Tensorflow.keras, Sequential Tensorflow.keras.layers, Dense, Dropout, LSTM, BatchNormalization Sklearn.model_selection, train_test_split Sklearn.preprocessing, StandardScaler Sklearn.metrics, confusion_matrix, classification_report

Comparisons

Traditional Machine Learning algorithms such as Logistic Regression and Random Forest are singular in their approach to tackling complex problems. These techniques need human intervention to break down numerous parts of the problem and combine these results in the final stages. Regardless of their efficacy and ease of implementation, a deep learning approach consists of multiple layers that work without human intervention to determine and identify patterns.

We believe that our selected model performed better than the others because neural networks are capable of not only identifying these hidden patterns, but also modeling and predicting these patterns. An LSTM model utilizes long-term memory in a more functional manner because it allows the model to train and learn on multiple parameters, making it powerful for forecasting purposes.

Limitations of the Study

The RNN-LSTM model has many parameters to account for. The optimal Dropout rate is difficult to pinpoint, which could cause the model to overfit. Additionally, finding the right activation is important because it determines which inputs are important and which inputs to ignore. Other considerations for improving our model include adjusting the nodes and hidden layers, the amount of units in the dense layer, and the number of epochs. In terms of improving model performance, we could increase the amount of units in the dense layer to 5 or 10. We could also opt to increase the number of epochs to 50 or even 100 to increase the validation accuracy while ensuring the training accuracy decreases. Although we had a lower training accuracy compared to the validation accuracy, the values were comparable, suggesting a potential overfitting issue.

Future Work & Conclusion

We started with four models and worked our way down to one: RNN-LSTM. We found that the model was able to classify with a high degree of accuracy, and that the previous four recessions were correctly labeled. With the addition of the 2022 data, we were able to produce a result that showed the country is currently in the midst of an economic recession.

We found that we can look to the yield curve to determine recessions, and that deep learning can be an immersive and eye-opening approach. Though we were limited to 2022, our call to action is for researchers and economists alike to study the yield curve and neural nets in hopes of being able to predict economic transgressions years into the future.

In closing, we achieved what we set out to do and in the meantime learned that the answer to many problems lies in deep learning. While we cannot fully place our faith in a model built in a few weeks, we can agree that current economic conditions are indicative of a recession.

References

Code: https://github.com/himabindu-thota/W207-AppliedML-final-project

1. Birken, Emily Guy. “Understanding the Yield Curve.” *Forbes*, Forbes Magazine, 12 July 2022, https://www.forbes.com/advisor/investing/yield-curve/.

2. Conerly, Bill. “The Yield Curve Predicts Recession-Sometimes, but Not Today.” *Forbes*, Forbes Magazine, 17 May 2022, https://www.forbes.com/sites/billconerly/2022/05/12/the-yield-curve-predicts-recession-sometimes-but-not-today/?sh=32df778a6544.

3. Q.ai - Powering a Personal Wealth Movement. “Recession vs. Bear Market in the S&P 500 and Dow Jones.” *Forbes*, Forbes Magazine, 17 June 2022, https://www.forbes.com/sites/qai/2022/06/15/recession-vs-bear-market-in-the-sp-500-and-dow-jones/?sh=3af4e1756a3e.

4. Reserve Bank of Australia. “Bonds and the Yield Curve: *Reserve Bank of Australia*”, Reserve Bank of Australia, 23 June 2021, https://www.rba.gov.au/education/resources/explainers/bonds-and-the-yield-curve.html.

5. Stojanovic, Dusan, and Mark D Vaughan. “The Yield Curve as a Forecasting Tool: St. Louis Fed.” *Saint Louis Fed Eagle*, Federal Reserve Bank of St. Louis, 9 Dec. 2021, https://www.stlouisfed.org/publications/regional-economist/october-1997/yielding-clues-about-recessions-the-yield-curve-as-a-forecasting-tool.

6. Torabi, Farnoosh. “Is the US Officially in a Recession? What to Know about Layoffs, Debt and Investing.” *CNET*, https://www.cnet.com/personal-finance/banking/is-the-us-officially-in-a-recession-what-to-know-about-layoffs-debt-and-investing/.