MGT 6203 Group Project Proposal

TEAM INFORMATION (1 point)

Team # 36 Members:

1. Kerstin Fontus; GT ID: 903943100
    - Educational Background: B.S. Mathematics and Scientific Computation – UC Davis
    - No analytic work background or major projects
2. Loraine Vizconde; GT ID: 903888947
    - Professional Background: Investment Operations Team Lead for an asset management firm
    - Education: B.S. Business Administration major in Economics from DLSU-D, Philippines
3. Edmund Dale; GT ID: 903755094
    - Professional Background: Solar energy researcher and materials scientist.
    - Education Background: BS in Physics from UCSD.
    - Analytics projects: Forecasting statewide renewable power resources and production.
4. Lukasz Sledz; GT ID: 903848861
    - Professional Background: Civil Engineer project manager in the transportation industry.
    - Education Background: BS in Civil Engineering from UIC.

OBJECTIVE/PROBLEM (5 points)

Project Title: A Modern Approach to Recession Forecasting

Background Information on chosen project topic:

When is the next recession? This is certainly the question asked after aggressive interest rate hikes and persistent inflation. This is the pattern that economists have seen for years, and it is also what was expected to happen. But what is a recession? A recession in the United States is defined as a significant decline in economic activity spread across the market lasting more than a few months, normally visible in real GDP, real income, employment, industrial production, and wholesale-retail sales (Per the San Francisco Federal Reserve, citing the National Bureau of Economic Research.)

Recession has always been painful to all sectors of the economy. Some view it as a necessary reset, but it still negatively affects both individuals and companies. To anticipate these challenges and better prepare for the future, it would be so valuable if recession could be predicted. This study aims to find certain factors that can aid in prediction.

Yield curve. According to the publication of the Federal Reserve Bank of New York, the yield curve is a valuable forecasting tool that is simple to use and significantly outperforms other financial and macroeconomic indicators in predicting recessions two to six quarters ahead.

(https://www.newyorkfed.org/medialibrary/media/research/current_issues/ci2-7.pdf)

The inverted yield curve has predicted recessions in the past, but obviously was not accurate for the short 2020 recession. The most popular opinion (38%) on the subject as found by the National Association of Business

Economists (NABE) was that inflation will decline without us sinking into a recession. The next popular (26%) was that there will be a recession in the next 1-1.5 years. Another 14% said low long-term premiums and no recession. Some have a theory that the inverted yield curve is no longer the best measure of an oncoming recession, thus, we will try to find other good or better factors for recession prediction.

(https://www.cnbc.com/2023/09/05/recession-indicator-may-be-broken-odometer-for-economy-says-expert.html)

Stock market index returns. The stock market is, without a doubt, a leading indicator of the performance of the economy. A well-performing economy produces a bull market where the prices of stocks continue to rise. A bear market is the opposite, where the prices tend to go downwards. This study will track the returns of the 3 indexes such as S&P500, Dow Jones, and Nasdaq to validate that the bearish market is a prelude to an impending recession.

(https://www.capitalgroup.com/advisor/insights/articles/guide-to-recessions.html)

M-Score. A recent paper finds that an aggregate M-score predicts recessions 5-8 quarters ahead of time, due to the fact that it predicts lower real investment 1-4 quarters ahead. M-score catches fraud in corporate earnings reports using 8 financial ratios. Based on the results of the paper, we hypothesize that this variable will be an important addition to our model. (https://publications.aaahq.org/accounting-review/article-abstract/98/5/129/10061/Aggregate-Financial-Misreporting-and-the)

Producer Price Index (PPI). In addition to the factors listed above, economists have studied many other macroeconomic factors associated with recessions. We will include PPI in our models because studies show that it is strongly correlated with recessions: Modeling and predicting U.S. recessions using machine learning techniques | Request PDF (researchgate.net)

Problem Statement (clear and concise statement explaining purpose of your analysis and investigation):

NABE is indicating that there are possibly better predictors for a recession than the yield curve. We are trying to find one first and foremost by looking at stock prices and returns, to see if that will yield accurate predictions. We will compare our results with analysis of more traditional macroeconomic factors. We will also build models with different combinations of the most significant predictors and try to determine if more modern variables like M-Score make better performing models.

State your Primary Research Question (RQ):

Can we predict the beginning and end of a recession using traditional indicators like yield curve, producer price index, and stock market index returns, as well as more modern predictors like Beneish M-score?

Add some possible Supporting Research Questions (2-4 RQs that support problem statement):

1. What are the most reliable macroeconomic indicators of a recession? Do any of them "lag" GDP in time?
2. Does including the first and second derivatives of predictors (for example the velocity and acceleration of index returns) improve model accuracy?
3. Does the performance of certain industries or companies have an impact in predicting recession?

4. Among our predictors, what variable is the most consistent in accurately predicting a recession?

Business Justification: (Why is this problem interesting to solve from a business viewpoint? Try to quantify the financial, marketing or operational aspects and implications of this problem, as if you were running a company, non-profit organization, city or government that is encountering this problem.)

Economic booms and busts affect every aspect of company strategy and investment decisions: from hiring, to asset management, to marketing, and more. Better tools for predicting recession would help companies across every industry make better decisions and save money.

For example, in commercial real-estate development, a fear of recession can cause investors to pull out of a project stalling it midway through completion. Knowledge of an upcoming recession would help developers avoid this kind of costly situation.

DATASET/PLAN FOR DATA (4 points)

Data Sources (links, attachments, etc.):

https://www.capitalgroup.com/advisor/insights/articles/guide-to-recessions.html

https://www.nber.org/research/data/us-business-cycle-expansions-and-contractions

https://www.newyorkfed.org/medialibrary/media/research/current_issues/ci2-7.pdf

Yahoo Finance API: yfinance · PyPI (For historical returns from S&P500, DOW, NASDAQ and other indexes)

FRED: https://fred.stlouisfed.org/ (For historical GDP, Yield Curve, and PPI data)

Compustat Financial via Wharton Research Data Services: https://wrds-www.wharton.upenn.edu/ (For M-score, beta, and other historical data from company financial statements)

Data Description (describe each of your data sources, include screenshots of a few rows of data):

- Compustat is a comprehensive market and corporate financial database published by Standard and Poor's, covering thousands of companies worldwide, with info dating as far back as 1950. The WRDS provides students a sample of their "As First Reported – Quarterly" table, which holds self-reported financial data from thousands of companies ranging from 1968 to 2013. The variables shown are used to calculate the Beneish M-Score. A more detailed description will be in the final report.

```
m_data_full.head(10)
```

Python

	gvkey	fyearq	datadate	ACTQ	ATQ	COGSQ	DLTTQ	DPQ	IBQ	LCTQ	OANCFY	PPENTQ	RECTQ	SALEQ	XSGAQ	conm	sic	tic
206	1115	1996	12/31/1995	117.105	167.961	33.890	0.000	0.198	3.541	78.620	-1.345	7.972	64.265	54.988	14.645	ADAC LABORATORIES	3844.0	ADAC
207	1115	1996	3/31/1996	117.985	168.735	36.124	0.000	0.198	3.938	75.525	-1.370	8.022	69.997	58.438	15.148	ADAC LABORATORIES	3844.0	ADAC
208	1115	1996	6/30/1996	121.417	173.127	38.300	0.000	0.198	4.353	77.236	0.640	8.107	71.379	62.434	16.291	ADAC LABORATORIES	3844.0	ADAC
209	1115	1996	9/30/1996	134.832	186.628	39.517	0.401	2.235	4.805	80.512	-7.631	8.393	80.654	64.925	16.957	ADAC LABORATORIES	3844.0	ADAC
213	1115	1997	9/30/1997	149.003	205.477	42.553	0.308	2.491	6.027	64.327	13.585	11.555	99.495	72.480	19.237	ADAC LABORATORIES	3844.0	ADAC

- Federal Reserve Economic Data (FRED) is an online database with economic time series data from national, international, public, and private sources. It is maintained by the Research Department at the Federal Reserve Bank of St. Louis. We are using it as a source for Historical GDP, the yield curve, and other economic indices.

- Screenshot of GDP, Yield curve, PPI, and velocity/acceleration data:

DATE	GDPC1	WPSID62	WPSID62 trend	WPSID62 velocity	WPSID62 acceleration	T10Y2Y	T10Y2Y trend	T10Y2Y velocity	T10Y2Y acceleration
6/26/1978						0.11	0.205333333	-0.004638889	0.000166667
6/27/1978						0.14	0.200722222	-0.004638889	-0.000138889
6/28/1978						0.13	0.196055556	-0.004916667	-0.000347222
6/29/1978						0.15	0.190888889	-0.005333333	-0.000208333
6/30/1978						0.14	0.185388889	-0.005333333	0.00025
7/1/1978	6644.754	74.2	74.16666667	-0.45	-0.125				
7/3/1978						0.15	0.180222222	-0.004833333	0.000402778
7/5/1978						0.18	0.175722222	-0.004527778	1.39E-05
7/6/1978						0.18	0.171166667	-0.004805556	-0.000375
7/7/1978						0.18	0.166111111	-0.005277778	-0.000430556
7/10/1978						0.18	0.160611111	-0.005666667	-0.00025
7/11/1978						0.2	0.154777778	-0.005777778	-1.39E-05

- We are also using yfinance API for data on index returns:

```
: print(sp500_monthly_return.head())

Date
1975-01-31 00:00:00-05:00         NaN
1975-02-28 00:00:00-04:00    0.059886
1975-03-31 00:00:00-04:00    0.021694
1975-04-30 00:00:00-04:00    0.047265
1975-05-31 00:00:00-04:00    0.044101
Freq: M, Name: Close, dtype: float64
```

Key Variables: (which ones will be considered independent and dependent? Are you going to create new variables? What variables do you hypothesize beforehand to be most important?)

We will mark the beginning of a recession when real GDP drops for 2 consecutive quarters, and we will mark the end of a recession when GDP rises for 2 consecutive quarters. Therefore, **GDP** will be considered our **dependent** variable, and our goal will be to accurately forecast GDP over different horizons. All other variables including **index returns, yield curve, PPI, M-score, and weighted average beta** will be considered **independent** variables. We will create variables for M-score and weighted average beta based on financial statement data from many companies in our index of choice (like S&P500). For both variables we will use market capitalization for the weighted average. We will also create variables based on the 1st and 2nd derivatives of all independent variables listed above and examine the predictive causality of these variables on GDP to see if they should be included in our models.

Based on our background research we believe that yield curve and M-Score will be the most significant variable used to predict GDP, but we will build models using multiple sets of variables to test this hypothesis.

APPROACH/METHODOLOGY (8 points)

Planned Approach (In paragraph(s), describe the approach you will take and what are the models you will try to use? Mention any data transformations that would need to happen. How do you plan to compare your models? How do you plan to train and optimize your model hyper-parameters?))

- For all daily and monthly data (like yield curve and PPI) we will use moving-average based seasonal decomposition to extract the "trend" component of these variables.
- M-Score data is not readily available, so it will be calculated quarterly from publicly available financial firm data.
- After feature engineering and data cleaning, we will build vector autoregressive models (VAR) to forecast GDP based on the time series predictor variables (described above) that "lag" GDP. We will use the following steps to build these VAR models:

1. Perform Augmented Dickey Fuller test (ADF) test on all variables to identify which time series are stationary and non-stationary.
2. Use differencing to transform all variables, then confirm stationarity with more ADF tests.
3. Use Granger causality tests to identify predictors that "lag" GDP. Include only these predictors in the initial VAR model.
4. Select order (p) of VAR model based on lowest AIC value.
5. Train VAR model of order (p). Experiment with different train test splits.
6. Check model residuals for serial correlation using Durbin Watson's statistic. If serial correlation is a problem, try making higher order VAR models or try adding more variables into the model.
7. Forecast GDP at different horizons.
8. Un-difference (inverse transform) the forecasted values to get the "real" forecast values in the correct units.
9. Plot the forecast versus actual data for different time horizons and evaluate the forecast accuracy using metrics such as: MSE, MAE, RMSE, MAPE, and R2.
10. Use Tableau to make an interactive visualization of historical data and forecasts.

Anticipated Conclusions/Hypothesis (what results do you expect, how will you approach lead you to determining the final conclusion of your analysis) Note: At the end of the project, you do not have to be correct or have acceptable accuracy, the purpose is to walk us through an analysis that gives the reader insight into the conclusion regarding your objective/problem statement

We all have a different hypothesis for our result:

Loraine: Stock market is an insignificant predictor to an economic recession. The volatility of the market gives an undetermined contribution to an economic downturn

Eddie: I suspect the model will accurately forecast GDP for short time horizons (1-2 quarters).

Kerstin: I believe the models containing M-Score and yield curve will end up being the most accurate models.

Lukasz: I suspect that analyzing the top indexes will be more conclusive to predicting recession.

What business decisions will be impacted by the results of your analysis? What could be some benefits?

Businesses could use this type of analysis to begin cutting costs sooner by slowing down production, conducting some layoffs, and liquidating assets early basically planning for being short on cash so they can avoid debt and bankruptcy. The analysis can also help businesses to manage and balance their portfolio in preparation for the economic downturn.

PROJECT TIMELINE/PLANNING (2 points)

Project Timeline/Mention key dates you hope to achieve certain milestones by:

1) Compile all the factors we want to consider and clean the data sets (mid-end October)
2) Finish building models (mid-end November)
3) Finish analysis/Report (end Nov-Dec 3)