Predicting Credit Card Defaults

Introduction –
The Project

- Economic downturns represent significant operating risk to the lending industry. They are also notoriously hard to predict
- Rather than attempting to predict economic downturns, this project seeks to predict the outcome that lenders care about: **Credit Card Default Rates**
- Publically available macroeconomic data is used to predict the future path of the aggregate credit card default rate, as recorded by the Federal Reserve

Sourcing the Data

- Nearly a hundred economic datasets were selected from the Federal Reserve Economic Data (FRED), a publically maintained repository for economic datasets.
- Variables from several categories were selected:
  - Aggregates designed to predict macroeconomic outcomes
  - Measures of employment
  - Metrics for aggregate and sector-specific economic production
  - Interest Rates
  - Stock Indices
  - Measures of Inflation
  - Measures of Consumer and Producer Sentiment
  - And other popular measures on FRED

Wrangling the Data

- The data required significant manipulation to process it into a usable form:
    - Daily and weekly data were resampled into monthly features, including average monthly day-over-day and week-over-week changes of each variable, the median value for each variable each month, and many more calculations
    - Quarterly and annual data were imputed to monthly data using a variety of methods
    - Month-over-month and year-over-year changes in each variable were added as separate features
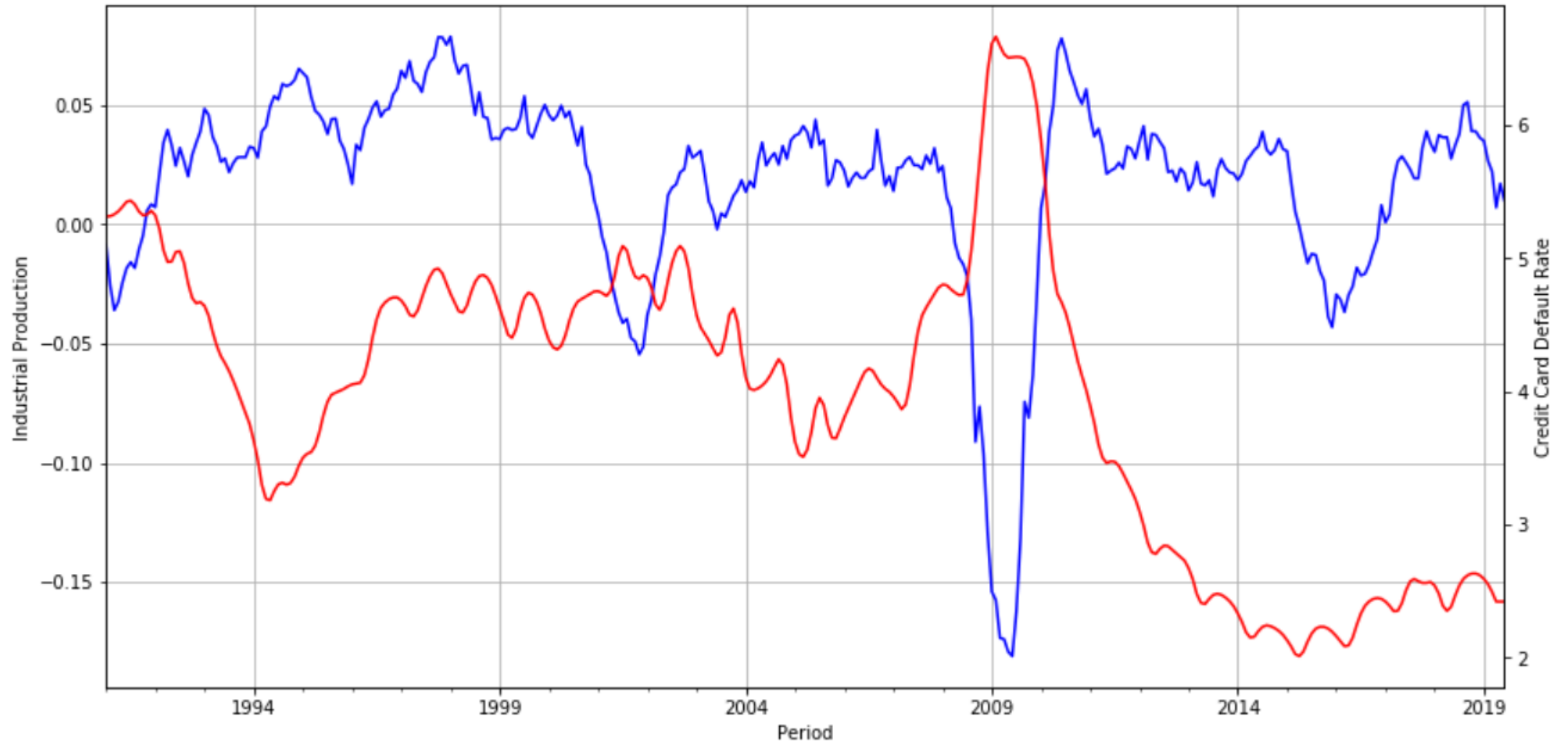
Exploratory Analysis

- Several manipulations were made to nearly 100 features, making exploration of each feature impractical
- Exploratory analysis was performed on some of the more promising variables. The analysis was designed to validate that:
  - Some variables track closely with the outcome variable, the Credit Card Default Rate. If no variables can be found that correlate strongly with the default rate, then it will be difficult for a model to predict the default rate.
  - At least some of the variables *lead* changes to the Credit Card Default Rate. The goal of the model is to *predict* defaults, which will be impossible if variables lag or directly overlap the default rate.

Exploratory Analysis – Industrial Production

- The Industrial Production Index measures all industrial production within the United States, including manufacturing, some natural resource production, and some utilities.
- Industrial production is generally known to track closely with economic output, as manufacturing is a large component of total economic activity. Industrial Production steadily increases over time as the economy grows, so the *year-over-year change* in industrial production is considered.
- The next slide shows the year-over-year change in industrial production graphed as a time series with the aggregate credit card default rate.
  - As expected, Industrial Production has a very strong looking inverse relationship with credit card defaults.
  - The variable looks as though it might begin to change before credit card defaults - for example, Industrial Production rises sharply in 1991, *after which* credit card defaults appear to fall.

Exploratory Analysis – Industrial Production
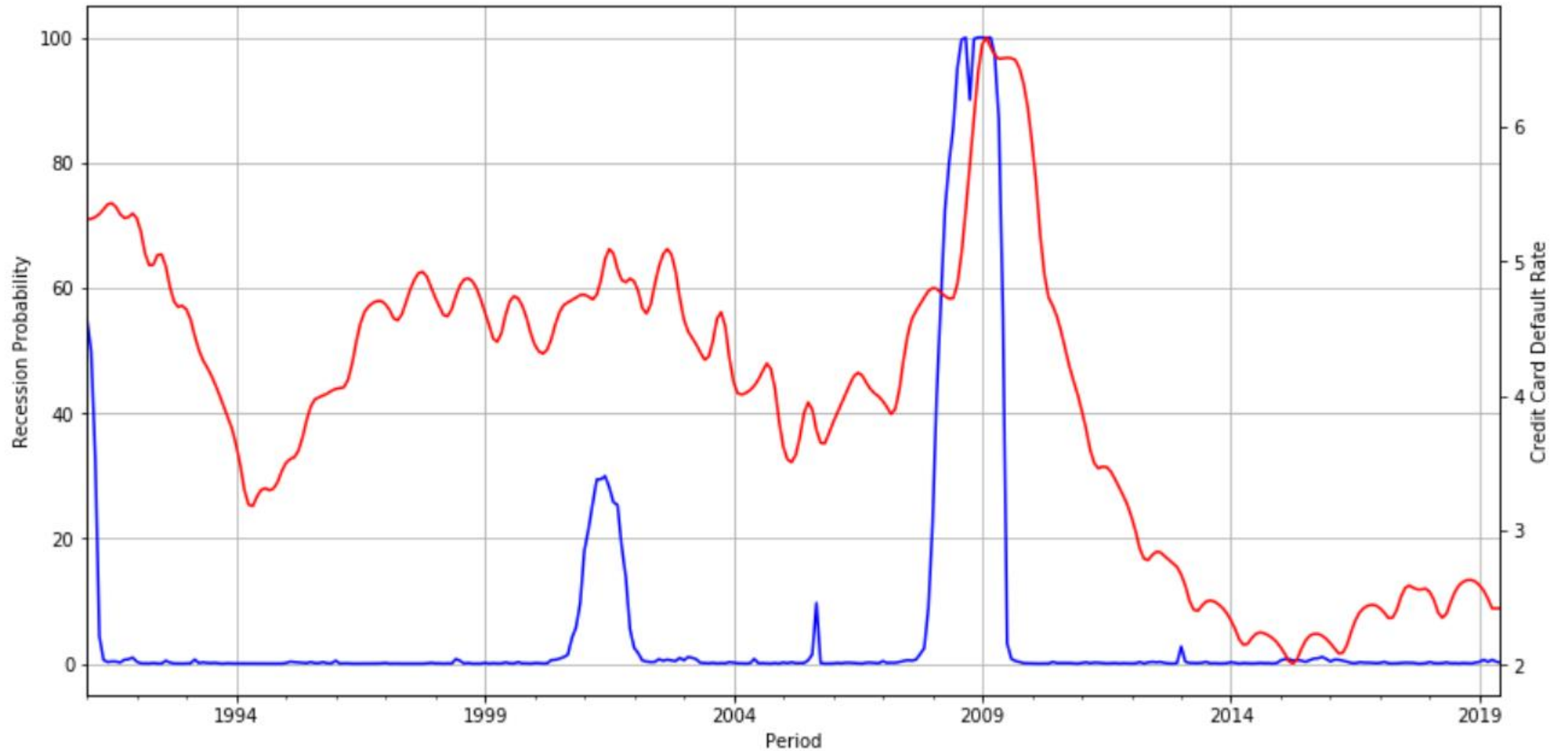
Exploratory Analysis –
Recession Probability

- The start date of a recession is often not identified until several months after its start. For this reason, there is value to predicting whether or not the economy is currently in recession.

- The Smoothed U.S. Recession Probabilities metric is a measure of the likelihood that a recession would occur during each month. As an economic recession typically severely increases credit card default rates, it is expected that a rise in the recession probability will be associated with an increase in defaults.

- The next slide shows the relationship between recession probabilities and the Credit Card Default Rate. The probability of a recession is typically zero, but when it increases it is often associated with an increase in the rate of credit card defaults, and it seems to be a very strongly leading indicator.
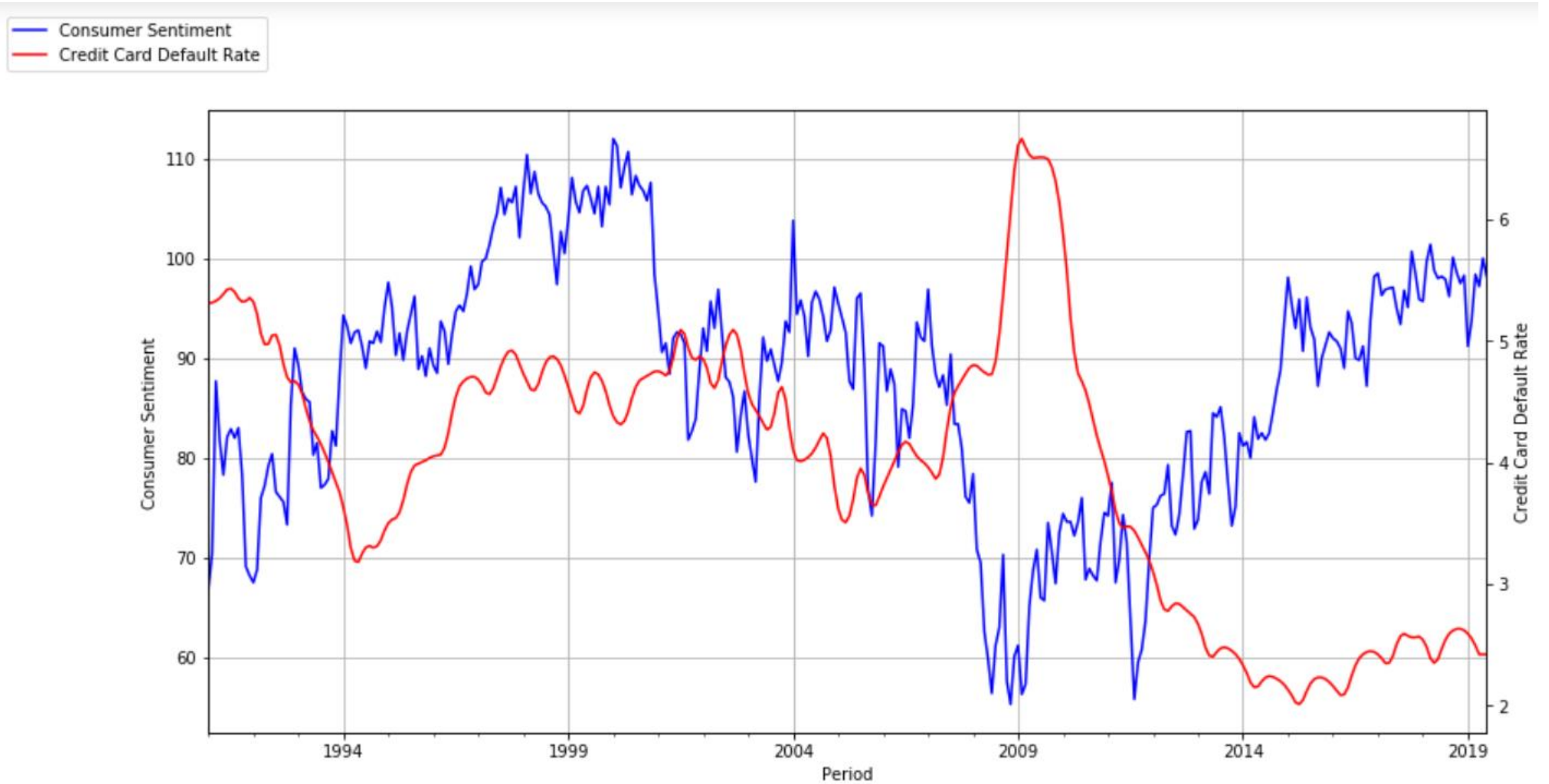
Exploratory Analysis – Recession Probability

Exploratory Analysis – Consumer Sentiment

- The University of Michigan conducts monthly surveys of consumers to evaluate how they perceive the health of the economy, the business environment, and their own personal finances.

- Results are aggregated into a single index that is designed to have a value of 100 during 1966Q1 - values greater than 100 represent sentiment that is better than this benchmark, and values less than 100 represent sentiment that is worse.

- The next slide shows consumer sentiment plotted with the credit card default rate. It is very common, as expected, for an increase in consumer sentiment to be associated with a decrease in the credit card default rate shortly after, which makes consumer sentiment a very good leading indicator.

Exploratory Analysis – Conclusion and Next Steps

- There appear to be a significant number of economic variables that closely predict the credit card default rate, which should provide sufficient information to train a Machine Learning model to predict defaults accurately.
- The next steps in this project are to draft multiple ML models using the already-pulled data and to evaluate how well they predict defaults
- Additional tweaks will be performed to iteratively improve predictions, including:
  - Incorporating more datasets
  - Refining and adding to the data engineering already performed
  - Trying different machine learning models and tuning hyperparameters for each

Thank You