Databusters Report - Forecasting Recessions

By Group 18 Xander, Ryan, Bryce, Yi Hong

1. Introduction

This report details the process of cleaning and selecting features from the FRED dataset to build time series forecasting models to predict economic recessions.

We decided to use quarterly data instead of monthly data because GDP is only available in the quarterly dataset. Furthermore, most variables present in the monthly dataset are also available in the quarterly dataset. By focusing on quarterly data, we ensure a more consistent and complete dataset for model training.

---

2. Exploratory Data Analysis (EDA) and Data Cleaning

2.1 Applying Transformations

Each variable underwent transformation based on its assigned transformation code (TCODE), as suggested by Stock and Watson's macroeconomic variable analysis (Stock & Watson, 2016). The transformations included differencing (first and second order), logarithmic transformations, and percentage change calculations. These steps ensure stationarity, which is crucial for effective time series forecasting.

2.2 Data Cleaning

We chose to include only variables considered in Stock and Watson's factor analysis, as these have demonstrated strong predictive power for economic trends. By filtering for variables with factor value 1, we significantly reduced the dataset's dimensionality from 245 variables to 125 variables, making further computations more efficient.

To handle missing data effectively, we first created a heatmap (Figure 1) to visualize missing values across the dataset. We noticed that several columns had many missing values because certain variables were only tracked from specific years onwards.

To handle the missing data, we only kept the columns with at least 75% of available data, and then proceeded to impute the remaining missing values using forward and backward fill.

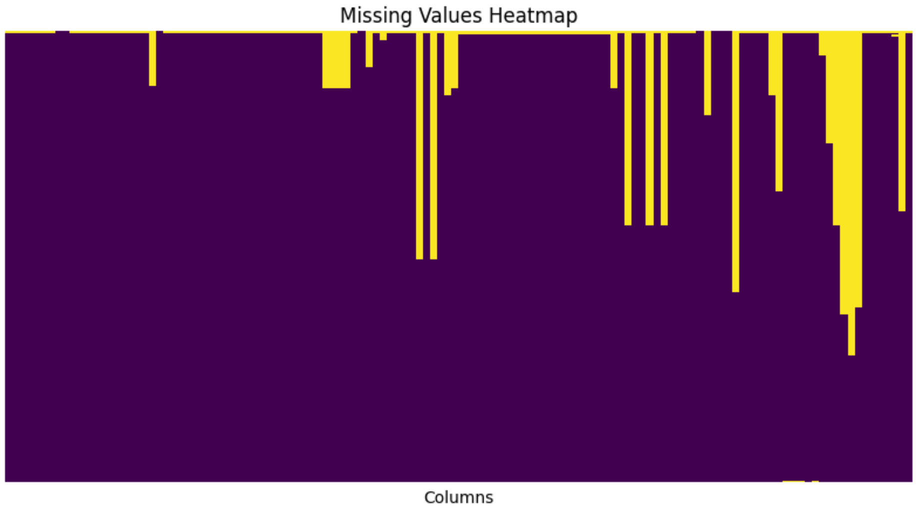

Figure 1. Heatmap of missing values for variables from 1955 Q1 to 2024 Q4

---

3. Feature Selection

3.1 Reducing Multicollinearity

For feature selection, a correlation matrix was calculated to identify highly correlated features. This step is critical to reducing multicollinearity, which can distort model predictions and lead to overfitting. A threshold of 0.85 was used to drop features with high correlation, ensuring only independent features were retained.

3.2 Feature Selection based on Domain Knowledge

Afterwards, 5 key features are identified based on domain knowledge and economic research.

According to FRED-QD (2010), the top five variables explaining total variation in their data were LNS14000025 (Unemployment Rate - 20+ Men), DMANEMP (Durable Goods Employment), LNS13023621 (Job Losers - Unemployment Level), USTPU (Trade, Transportation & Utilities Employment), and IPBUSEQ (Industrial Production: Business Equipment). Four of these relate to the labor market, while one reflects industrial production. LNS14000025 is inversely related to GDP growth, aligning with Okun's Law, where a 1% rise in unemployment leads to a 2–3% GDP decline. DMANEMP is procyclical, rising during economic expansions and falling during downturns, as durable goods manufacturing is sensitive to consumer demand and business investment. Similarly, LNS13023621 is inversely correlated with GDP, as increased job losses reduce consumer spending and economic growth. USTPU employment is also tied to GDP but lags by two to three quarters due to hiring delays. IPBUSEQ, on the other hand, serves as a leading indicator, with rising production signaling stronger business investment and GDP growth.

These 5 features were added back into the dataset. A check for missing values was performed to verify data integrity. Finally, the data was split into a 70/30 ratio for training and testing, which is standard practice to ensure that models are tested on unseen data for robust evaluation.

---

4. Time Series Forecasting Models

We implemented two forecasting models: ARIMA, and VAR. Each model was chosen based on its strengths in capturing different aspects of economic trends.

4.1 ARIMA Model

The ARIMA (Auto-Regressive Integrated Moving Average) model was chosen because it accounts for trends in the data. Although the tcode was specified, we ran it through this model in the event further differencing is better suited. From the feature selection, the best (p, d, q) parameters with the lowest AIC value is (2, 0, 2). AIC was used as it is better suited to minimise prediction error, and when the true model is too complex to estimate parametrically. The advantage of ARIMA is its flexibility in capturing both short-term and long-term dependencies. However, its drawback is that selecting the optimal (p, d, q) parameters can be computationally expensive and may require significant tuning.

4.2 VAR Model

The Vector Autoregression (VAR) model was chosen because it is well-suited for capturing the interdependencies between multiple time series variables. In this case, the dataset consists of multiple macroeconomic indicators (e.g., GDP, inflation, employment) that may influence each other over time. Unlike univariate models such as ARIMA, VAR can handle multivariate data, making it a powerful tool for forecasting when variables are believed to have dynamic relationships. However, a limitation is that VAR models can become overly complex with too many lags and variables, leading to overfitting, especially in small datasets.

Using AIC to evaluate model performance

We chose to use the Akaike Information Criterion (AIC) over the Bayesian Information Criterion (BIC) to evaluate our model’s performance due to our focus on minimizing predictive error rather than strictly identifying the true model. While BIC is asymptotically consistent, meaning it will select the true model if it exists within the candidate set as the sample size increases, AIC is more efficient when the true model is not included in the candidate models. AIC asymptotically selects the model that minimizes mean squared prediction error, which aligns with our goal of improving prediction accuracy for economic contractions. Additionally, AIC’s minimax property, which minimizes the maximum possible risk in finite sample sizes, provides better protection against overfitting in scenarios where sample sizes are limited. Given the complexity and potential non-nested nature of our models, AIC was better suited to our needs for flexibility and accuracy in forecasting U.S. economic downturns.

---

5. Model Evaluation

To assess model performance, we computed directional accuracy, which measures how well the predicted directional changes align with actual movements. This metric is particularly relevant for economic forecasting, where predicting the correct direction of change is often more valuable than predicting exact values. The directional accuracy of the ARIMA model was about 0.40, whereas the VAR model was about 0.50.

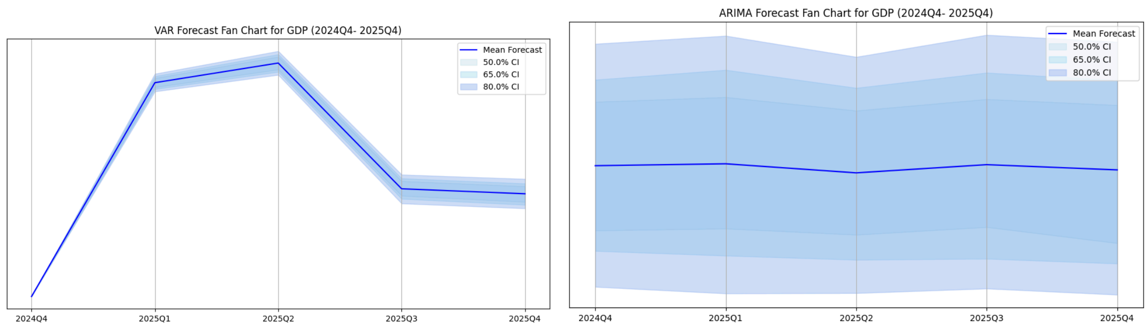

Figure 2. Line graph of forecasted GDP against Quarters 2024Q4 -2025Q4

---

6. Conclusion

This report outlines a structured approach to data cleaning, transformation, and feature selection for economic forecasting. The choice of quarterly data ensures consistency with GDP reporting, and the selection of ARIMA, and VAR models allows for a comprehensive analysis of economic trends. Future steps include refining model selection, incorporating additional economic indicators, and experimenting with alternative forecasting techniques to improve predictive accuracy.

References:

Bhosale, R. (2016) ‘Relationship between unemployment and economic indicators’, *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.3133472.

Dossche, Maarten & Saiz, Lorena (2018) ‘Consumption of durable goods in the ongoing economic expansion’, Economic Bulletin Boxes, European Central Bank, vol. 1.

Stock, J.H. and Watson, M.W. (2016) ‘Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in Macroeconomics’, *Handbook of Macroeconomics*, pp. 415–525. doi:10.1016/bs.hesmac.2016.04.002.

Sánchez, J.M. and Liborio, C.S. (2012) ‘The relationships among changes in GDP, employment, and unemployment: This time, it’s different’, *Economic Synopses*, 2012(13). doi:10.20955/es.2012.13.

Van Leemput, E. and Ferraro, J.K. (2019) ‘Long-run effects on Chinese GDP from u.s.-china tariff hikes’, *FEDS Notes*, 2019.0(2382). doi:10.17016/2380-7172.2382.

Vrieze, S.I. (2012) ‘Model selection and psychological theory: A discussion of the differences between the Akaike Information Criterion (AIC) and the Bayesian information criterion (BIC).’, *Psychological Methods*, 17(2), pp. 228–243. doi:10.1037/a0027127.