DATABUSTERS Challenge 2025: (Team 32)

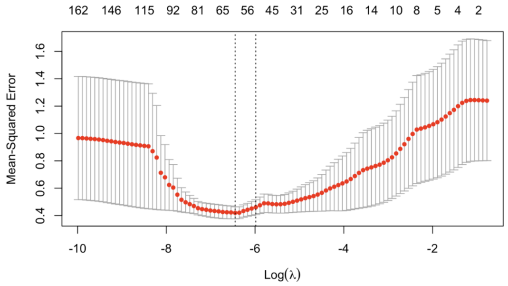
Predictor Selection

Note: Economic contractions in this report are defined as a quarter of negative real GDP growth.

Model Feature Selection for ARIMAX: A machine learning approach to complement our analysis.

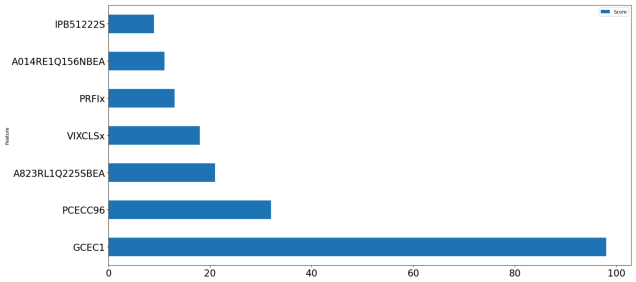
We created two regression models with feature importance capabilities so that we can statistically find the top features to feed into our prediction model. This is due to the nature of most time-series machine learning models (including ARIMAX) using all exogenous features that the user provides, and with the vast data we have for each quarter this would likely cause overfitting or multicollinearity. These regression models attempt to find the best predictors of GDP change, a quantitative variable, so that we can predict what the GDP change will be over the next four quarters.

1. LASSOCV



LASSO regression identifies key features by emphasizing linear relationships between predictors and the target variable. It applies internal cross-validation to select the optimal regularization strength, minimizing the mean squared error. To apply regression, we used GDP percent change, which we defined as the difference between the current year GDPC1 and the previous year GDPC1, divided by previous year GDPC1 and multiplied by 100 (to make it a percentage). The accompanying graph shows the tradeoff between minimizing MSE and non-zero features (on top) while creating the model. Inventory to sales ratio and help wanted/# unemployed were the top predictors in this model of relative GDP change. The following predictors were particularly important: A014RE1Q156NBEA, IPDCONGD, IPB51110SQ, UEMPLT5, UEMP15T26, HOUSTMW, BAA10YM, VIXCLSx, AAAFFM and TNWMVBSNNCBx

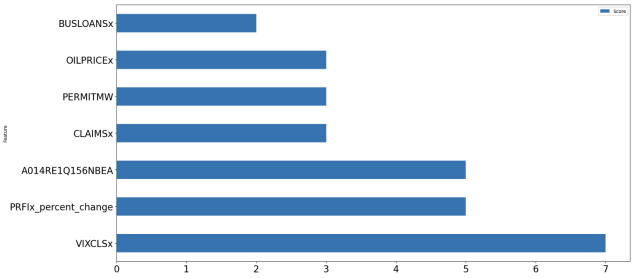
2. XGBoost



XGBoostRegressor identifies complex, nonlinear relationships between features and the target variable. The most influential features selected align with GDP components—consumption, investment, government spending, and net exports—including personal consumption expenditures, private domestic investment, and private fixed investments. Additionally, compensation per hour and petroleum prices emerged as key predictors, possibly serving as broader indicators of economic conditions. The accompanying graph is displaying the top features (larger = more important).

We applied the same concepts to find the best features that predict contractions. For this, we used XGBoost Classifier.

3. XGBoost Classifier



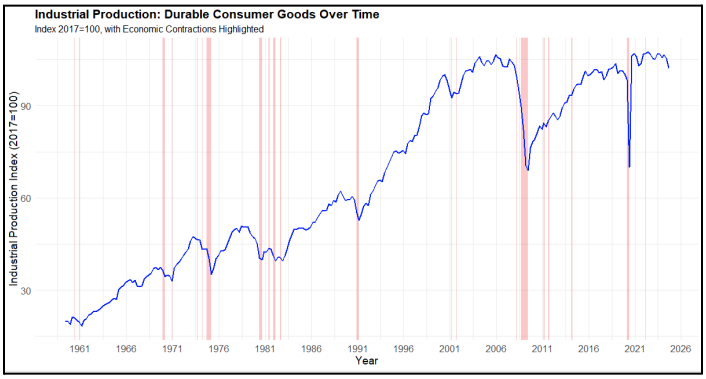
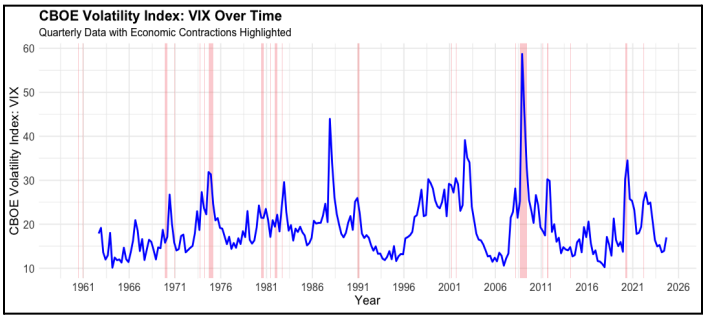
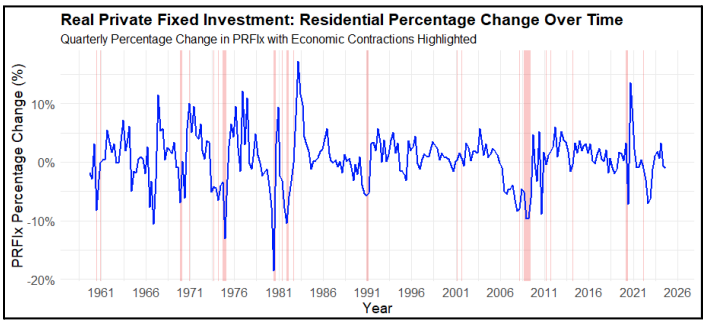
XGBoostClassifier, which predicts contraction likelihood, highlighted a different but complementary set of top features, including the volatility index (VIXCLSx), unemployment insurance claims (CLAIMSx), and building permits in the American midwest (PERMITMW). The accompanying graph is displaying the top features.

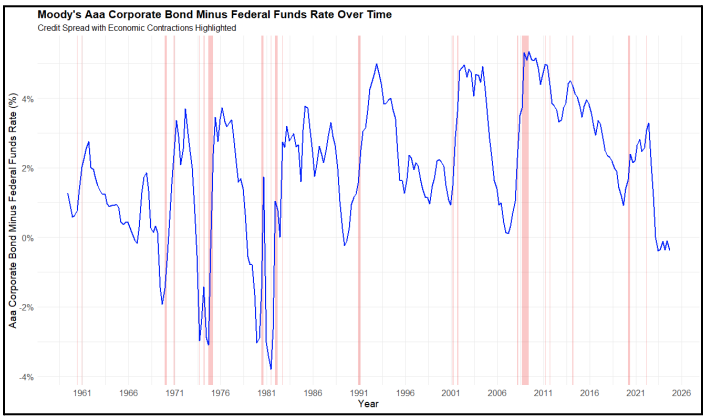
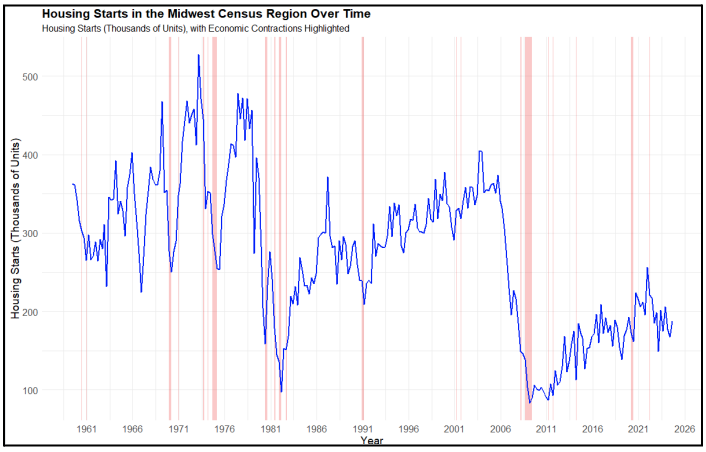
Exploratory Data Analysis

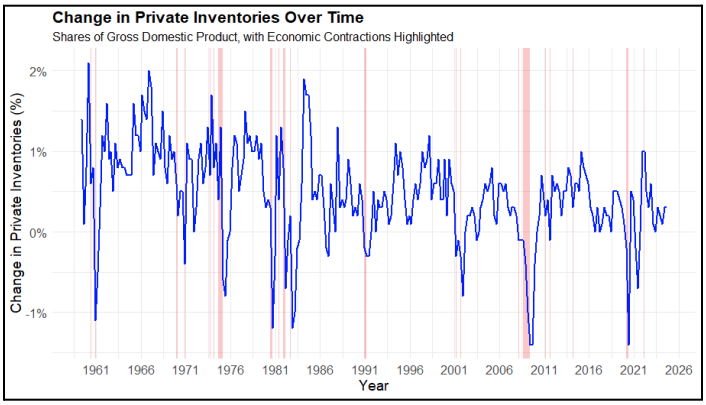
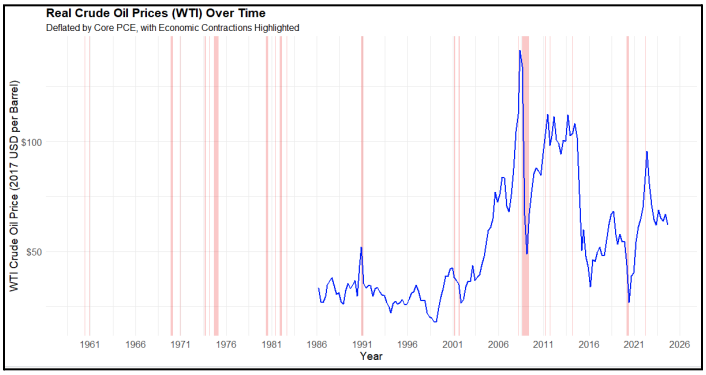
On top of computationally selecting features, we did research as to common forward indicators of contractions, and performed analysis on these features. Visually, IPBUSEQ, IPCONGD, and PRFIx stood out as potential features we would want to include. We will go over the analysis in the “Key Predictors Analysis” section. While originally not selected by our models, we engineered new features such as PRFIx_percent_change based on visual trends (in this case, we noticed sharp declines of PRFIx during contractions). Other engineered features fell just outside of the top seven shown in terms of importance and were not used in ARIMAX.

Key Predictors Analysis

For all plots under visualisation and analysis, the red regions represent periods of economic contractions. Detailed codes on how we obtained the charts could be found in our GitHub Repository.

Indicator	Economic significance	Visualisation and Analysis
IPDCONGD (Industrial Production: Durable Consumer Goods (Index 2017=100))	Consumer goods are products purchased by individuals for personal use, also referred to as final or retail goods. These goods represent the final stage of production and manufacturing. Examples of them include clothing and food. (Source:https://www.investopedia.com/terms/c/consumer-goods.asp) Consumer spending is the biggest factor influencing the US Economy. For example, according to the U.S. Bureau of Economic Analysis, in 2024's Q3, personal consumption expenditures accounted for nearly 68% of the U.S. GDP. This highlights the significant role consumer goods play in influencing economic growth or contraction. (Source: https://www.usbank.com/investing/financial-perspectives/market-news/consumer-spending.html)	 As seen, most economic contractions occur after a decline in IPDCONGD. Sharper drops are generally followed by longer periods of economic contraction.
VIXCLSx (CBOE Volatility Index: VIX)	The CBOE Volatility Index is a forward-looking index formed using the implied volatilities of the S&P500 Index. It measures the market expectations in the next 30 days using options from the S&P500 Index. Source: https://www.investopedia.com/terms/v/vix.asp An increase in the VIX index indicates heightened uncertainty, typically driven by market volatility, geopolitical tensions, or economic disruptions. Elevated uncertainty can reduce consumer spending and business investment, potentially slowing GDP growth. Source: (https://www.fidelity.com.sg/beginners/what-is-volatility/volatility-index)	 The chart plots the CBOE VIX against time, with red regions representing periods of economic contractions. The higher VIX indicates greater volatility and uncertainty in the market, leading to greater risk aversion in investors, leading to drop in investment expenditure and hence GDP growth. This is seen from the graph with peaks in VIX generally preceding contractions. Based on last 74 quarters, it's clear that higher CBOE VIX is correlated to contractions, with it peaking during 4/5 last contractions
PRIFx (Real Private Fixed Investment: Residential (PRFIx, deflated using its own price index, Billions of Chained 2017 Dollars) by Percentage Change	Real Private Fixed Investment: Residential (PRFIx) is an important economic indicator as residential investment typically declines before economic contractions. This is because residential investment is highly sensitive to interest rates, consumer confidence and credit conditions. Historically, declines in PRFIx have preceded recessions as higher mortgage rates, tightening credit and declining home demand reduce residential construction activity. PRFIx is a key component of gross fixed capital formation	

	and its fluctuations reflect the overall health of the housing sector. The housing sector being a major driver of employment, construction and consumer spending. (Source:https://www.bea.gov/resources/methodologies/nipa-handbook/pdf/chapter-06.pdf)	As seen, economic contractions usually occur after a decline in PRIFx. Drops over longer periods (shown by negative percentage change over longer periods) are generally followed by longer periods of economic contraction.
AAAFFM (Moody's Seasoned Aaa Corporate Bond Minus Federal Funds Rate)	The AAAFFM is a measure of the premium that top-rated corporate borrowers pay over the risk-free overnight lending rate set by the Federal Reserve. A higher Aaa-FFR spread suggests that corporate borrowing costs are rising relative to the risk-free rate, potentially due to increased credit risk perception or tighter monetary policy. This discourages business investment, leading to lower capital expenditure (CapEx), reduced hiring, and slower GDP growth.	 As seen, sharp drops in AAAFFM usually precede economic contractions. In particular, drops in AAAFFM into negative percents precedes economic contractions quite consistently.
UNEMPLT5 (Number of Civilians Unemployed - Less Than 5 Weeks (Thousands of Persons))	The UNEMPLT5 indicates the number of civilians that are unemployed for less than 5 weeks. It measures the number of people who have recently lost their jobs. An increase in this number reflects potential economic disruptions such as layoffs, which could be linked to cyclical downturns or disruptions leading to layoffs. A sharp increase in this number can signal an upcoming contraction as a leading indicator because it reflects a quick rise in layoffs — often associated with declining business confidence. Hence, it may predict a slowing economy.	 As seen in the graph, sharp rises in the number of unemployed civilians precede most of the economic contractions (highlighted in red).
HOUSTMW (Housing Starts in Midwest Census Region (Thousands of Units))	New housing starts fall under big-ticket capital purchases and is hence an indicator of change in consumption spending and consequently GDP change. When housing starts numbers change, it suggests strong demand, consumer confidence and economic expansion. On the contrary, when housing demand falls, it may indicate falling consumer demand and weakening economic conditions. This number also has implications on the labour market and hence overall consumer spending through loss of income. Midwest Census region was also the most affordable region in US with high demand for properties in that region.	

	This combination of factors and reasons explain why HOUSTMW is an important factor in determining real GDP in the US. (Source: https://finance.yahoo.com/news/why-the-midwest-remains-one-of-the-only-affordable-places-for-many-homebuyers-185613468.html)	As seen, sharp declines in Housing Starts usually precedes economic contractions. Sharper and longer drops in Housing Starts tend to be followed by longer periods of economic contractions.
A014RE1Q156NBEA (Shares of gross domestic product: Gross private domestic investment: Change in private inventories (Percent))	Inventory investment being a very volatile component of GDP, gives it an important role in short run variations in GDP growth. The change in private inventories plays a critical role in business cycles and has very strong predictive power for economic downturns. Businesses will adjust inventories based on expected demand. When demand weakens, businesses tend to reduce inventory accumulation in order to avoid overstocking. If the demand is weak for an extended period, they might draw down inventories, which cause this metric to turn negative. In the GDP formula of GDP = C + I + G + (X - M), I which represents investments include these private inventories. Thus a sharp decline or negative values in this metric can drag down GDP growth, making an economic contraction more likely. (Source: https://www.bea.gov/system/files/2019-12/Chapter-7.pdf)	 As seen, drops in Change in Private Inventories tend to be followed by economic contractions. In particular, drops to negative percentage values precedes economic contractions quite consistently.
OILPRICEx (Real Crude Oil Prices: West Texas Intermediate (WTI) - Cushing, Oklahoma (2017 Dollars per Barrel), deflated by Core PCE)	Oil is a key resource in many industries, ranging from transportation, manufacturing to energy production. It is a basic material used in many businesses. Higher fuel and energy costs also reduce disposable income, as consumers have to spend more on necessities like gas and electricity, leaving less for discretionary spending like entertainment, food and travelling, slowing economic growth. Source: (https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp362.pdf)	 As seen above, generally, when there are large increases in the price of Real Crude Oil Prices, an economic contraction follows. This could also be due to inflationary pressures. As seen from the long contraction period around the years of 2008-2009, large spikes in oil prices can also be seen to precede longer periods of economic contraction.

Forecasting

Overview

We used an AutoRegressive Integrated Moving Average with Exogenous Variables (ARIMAX) model to forecast GDP growth for the next four quarters (Q1 2025 - Q4 2025). Our exogenous variables chosen were due to economic relevance to GDP growth (As concluded from our Lasso, XGB Regressor and Classifier models as well as the analysis above). These included the following:

A014RE1Q156NBEA	IPDCONGD	IPB51110SQ	UEMPLT5	UEMP15T26	HOUSTMW	BAA10YM
VIXCLSx	AAAFFM	TNWMVBSNNCBx	CLAIMSx	PERMITMW	OILPRICEx	PRFIx_percent_change

Since GDP growth is sequential in nature, ARIMA-based models are ideal for capturing past trends. ARIMAX allowed us to incorporate the above economic indicators in our forecasting.

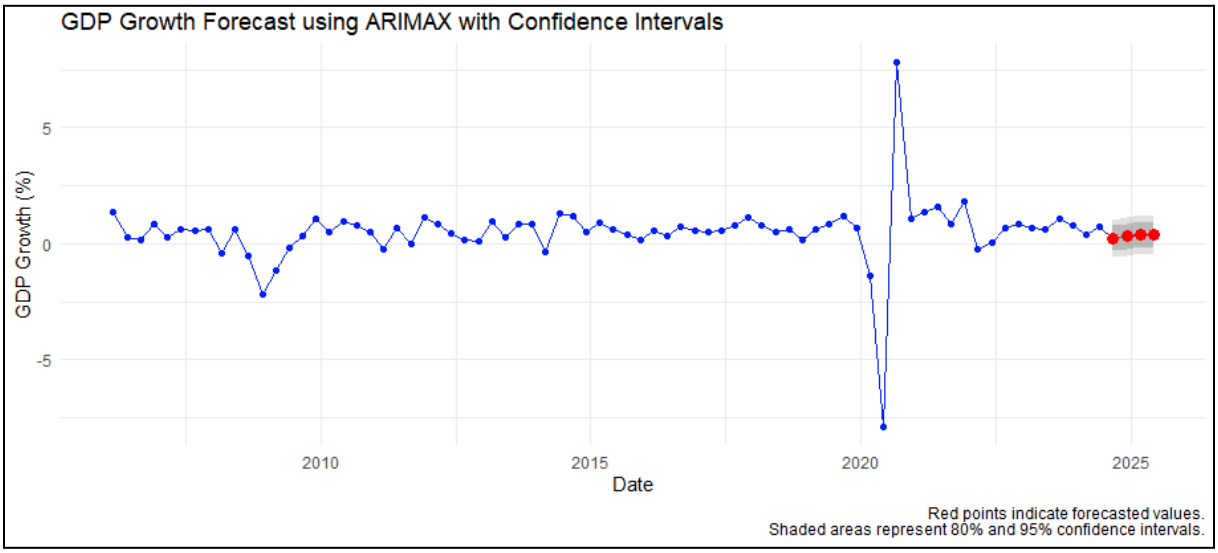
Testing for the stationarity of GDP Growth

```
        Augmented Dickey-Fuller Test

data:  gdp_ts
Dickey-Fuller = -6.5922, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
```

We conducted an Augmented Dickey-Fuller (ADF) test, which confirmed that GDP growth is stationary (p-value < 0.05). However, even though GDP growth is found to be stationary, that does not mean seasonality is absent. When we set seasonal = TRUE in ARIMAX, we may align seasonal effects in GDP growth with seasonal patterns in predictors, improving how the model interprets macroeconomic factors over time. Hence we made the decision to set seasonal = TRUE.

Forecast




```
> print(forecast_output)
  Quarter Predicted_GDP_Growth   Lower_80  Upper_80   Lower_95 Upper_95
1 Q1 2025            0.2418548 -0.2843332 0.7680429 -0.5628804 1.046590
2 Q2 2025            0.3070016 -0.2300155 0.8440186 -0.5142952 1.128298
3 Q3 2025            0.4097709 -0.1276918 0.9472336 -0.4122073 1.231749
4 Q4 2025            0.3842151 -0.1532661 0.9216963 -0.4377914 1.206222
```

As seen, all predicted GDP growth values are positive, which suggests no economic contraction at any of the forecast horizons. It also shows 80% confidence interval (gray region) and 95% confidence interval (wider gray region) encompassing negative values.

Conclusion: The ARIMAX model forecasts positive GDP growth for all future quarters, meaning it does not indicate an economic contraction in Q1 2025, Q2 2025 or Q4 2025. However, it should be noted that the confidence intervals indicate a moderate risk for economic contractions during all forecast horizons as the intervals dip into negative values, particularly during Q1 2025 and Q2 2025. An interesting note is that the forecasted GDP growth in Q4 is lower than that of Q3, indicating possible slowing of growth in the fourth quarter. While the forecast shows positive growth, macroeconomic factors like inflation, interest rates and geopolitical risks could alter actual GDP growth outcomes.