KPMG

S&P 500 Earnings Growth Forecast Models

Columbia University
MA, Quantitative Methods in the Social Sciences

Taotao Jiang, Albert Li, Peishan Li, Christina Lv, Jinghan (Katherine) Ma, Kushal Wijesundara, Michelle A. Zee

08/31/2021

The Team

Taotao Jiang

Healthcare Sector

tj2441@columbia.edu

Albert Li

Short-Term Overall Model

jl5813@columbia.edu

Peishan Li

Short-Term Overall Model
Information Technology Sector

pl2772@columbia.edu

Christina Lv

Recession Probability Prediction
Financial Sector

jl5727@columbia.edu

Jinghan (Katherine) Ma

Short-Term Overall Model
Telecommunications Sector

jm5223@columbia.edu

Kushal Wijesundara

Long-Term Overall Model
Ensembled Overall Results

kcw2144@columbia.edu

Michelle A. Zee

Fed Funds Rate Prediction
Consumer Discretionary Sector

maz2136@columbia.edu

Executive Summary

Context

In recent years, the business world has been experiencing high uncertainties due to rapid technological growth and business cycle fluctuations. As a result, it is crucial for companies to predict the future performance for the overall market and various sectors.

Objective

To assist clients with forward-looking market predictions, we build multiple machine learning algorithms incorporating **2 unique features** that forecast short- and mid-term S&P 500 overall and sector-specific earnings performance.

Predictive Model

1	2	3	4	5	6
Facebook Prophet	XG Boost	LDA Allocation	Support Vector Machines	SARIMAX	Random Forest Regression

Agenda

- Introduction
- S&P 500 Overall Models
- S&P 500 Sector-Specific Models
- Q&A

Objective

Use traditional **economic indicators** and **engineered features** to provide **1 - 18 month forecasts** of **earnings growth** of the **overall S&P 500** and the **S&P 500 Sector-Specific Indices**.

Used for:

- Inform businesses on the future market conditions
- Contribute to a firm's operational, capital expenditure, and financial planning

Target Variable: Monthly Normalized Earnings Growth

Earnings = Monthly Index Price ÷ LTM P/E Ratio

Target Variable = (Earnings$_t$ - Earnings$_{t-1}$) / Earnings$_{t-1}$ - 10-Year Treasury Rate

S&P 500 Overall Data: 1969 - today

S&P 500 Sector-Specific Data: 2001 - today

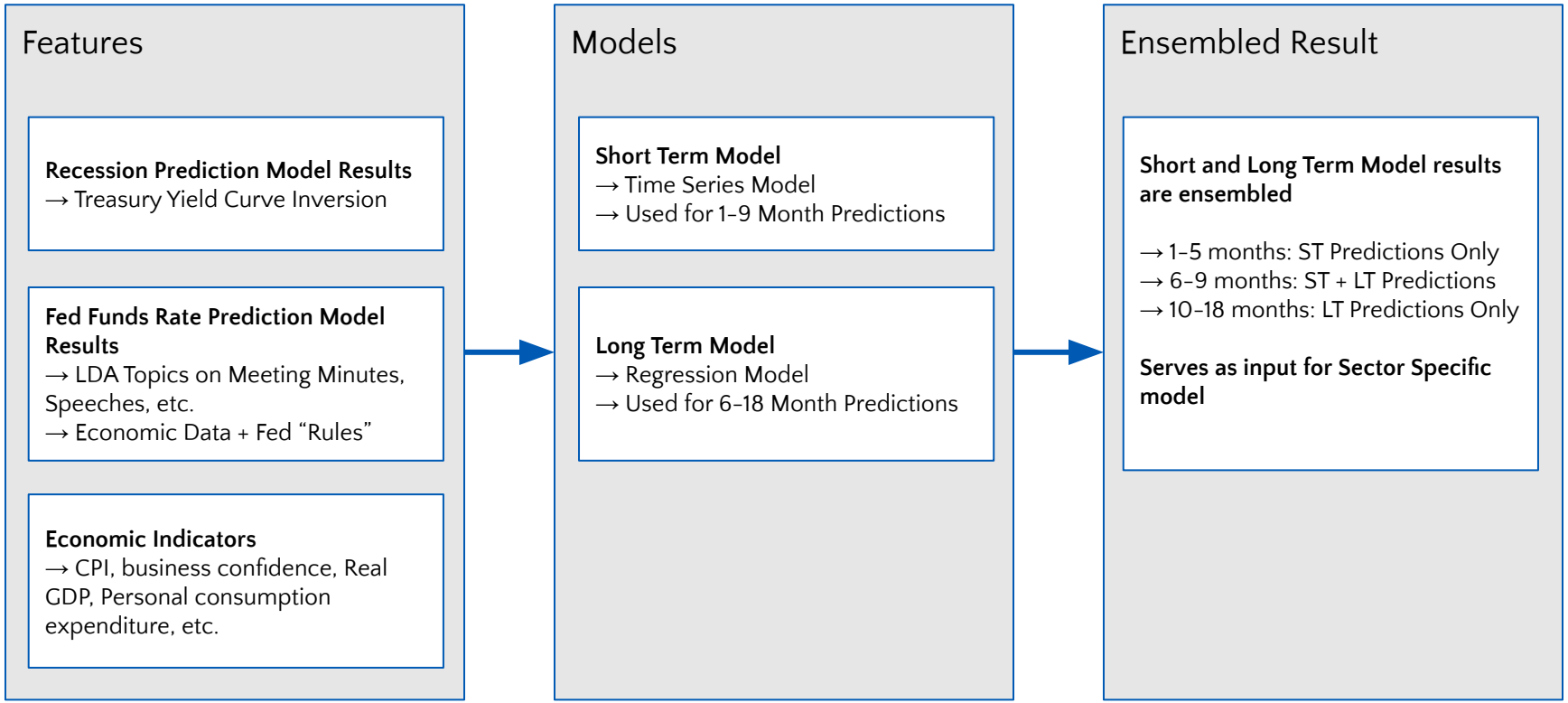
S&P 500 Overall Models

Model Journey

Earnings Growth Forecasting Approach Criteria

- Overall economy & macroeconomic trends
- Capture inflexion points:
    - Fed reserve monetary policy: interest rates & QE
    - Recession prediction: business cycle
- Time: important to capture the 1970s and 80s *hyperinflation period

S&P 500 Overall Index Model Overview



Predicting Recession Probability

Why is "Recession Prediction" important to the Overall Market Model?

- Identify where we are on the business cycle by capturing downward shocks
- Signal near-term recession potentially in the next 12 months

How is the Recession Model relevant to our client?
It helps to send early warning signals and improve earnings management.

Model Objective: Take the difference between 10-year and 3-month Treasury rates (also defined as the "**term spread**"), as well as other economic indicators to calculate recession probability in 12-month ahead horizon

Positive Term Spread	When the yield on long-term US Treasury bonds is higher than the yield on on short-term Treasury bills.
Negative Term Spread	**Investors holding short-term treasury bonds get paid more than those in long-term ones.**

Inversion of the Yield Curve indicates a potential economic downturn

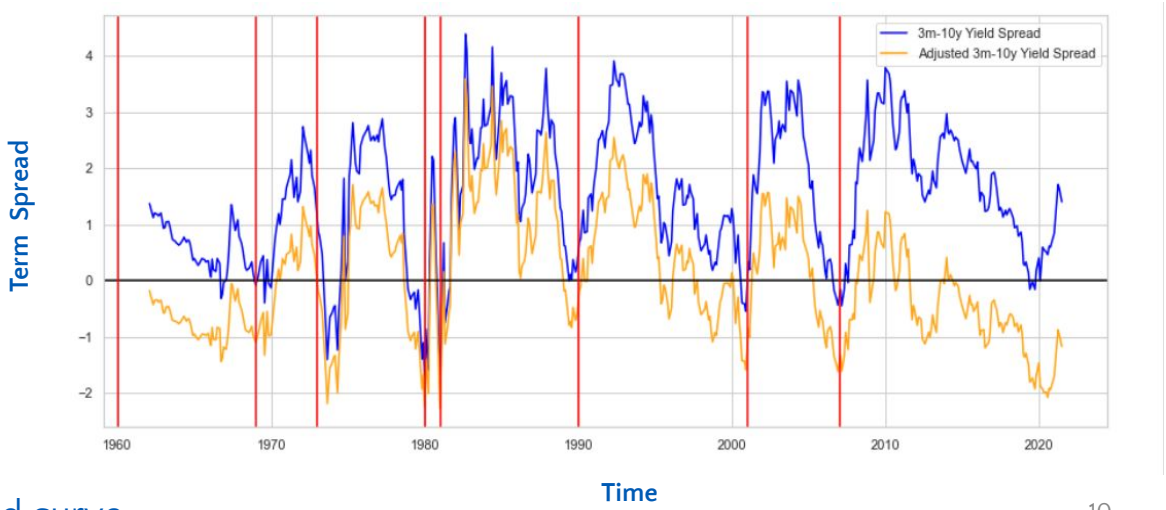
Adjusting the Yield Curve to Amplify Recession Signals

An increasingly noisy signal?
The Fed have pushed the entire US treasury yield curve so far down that an inversion has become more difficult to achieve than in the past. **If we take the yield spread at face value, we will forecast a low probability of a recession as long as money market rates remain low.**

Adjusting the Yield Spread

$$ASP_t = \varepsilon_t = SP_t - 2.616 + 0.808 \ln(STR_t + 1)$$

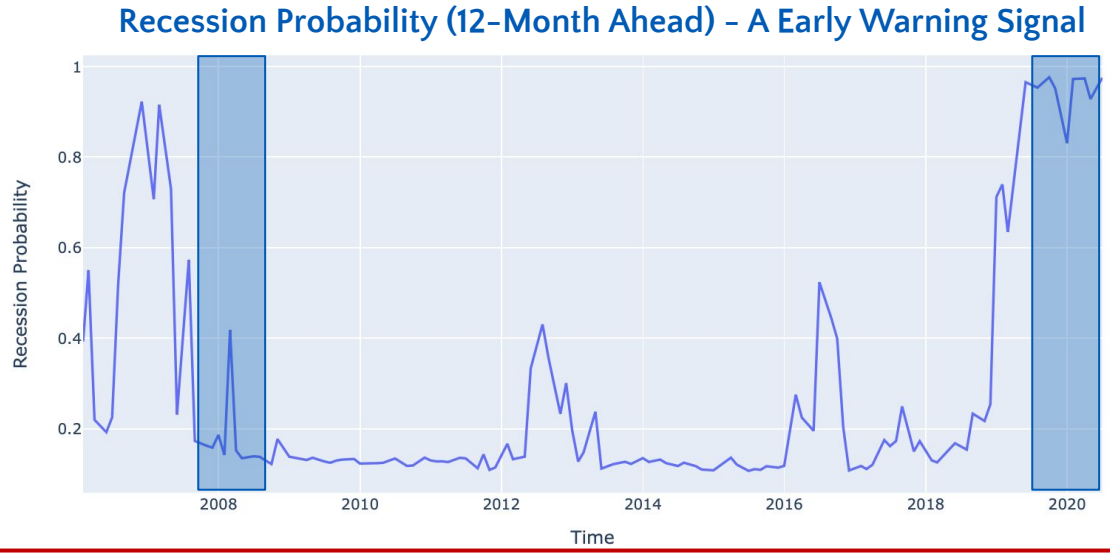
Adjusted 3m-10Y Yield Spread (ASP) vs. 3m-10Y Yield Spread (SP)



Reference: robabank – adjusted yield curve

12-Month Ahead Predictions

Dependent Variable (12-Month Ahead Prediction):
NBERt,t+12, equals one if there is an NBER recession starting at any time in the 12 months that follow the observed independent variables, and zero otherwise.



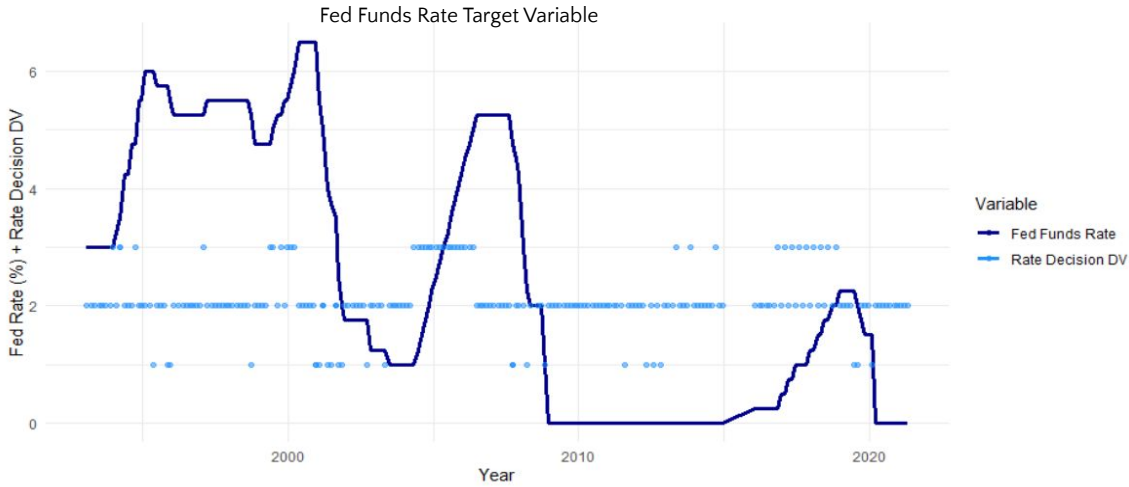
The adjusted yield curve is increasingly emitting warning signals of a recession. For instance, recession probability on February 2021 represents the likelihood of economy being in a recession in February 2022.

Fed Funds Rate

- Model predicts future monetary policy, which affects overall interest rates and liquidity

- Target variable: categorical variable signaling raise, hold, or lower of FFR
- Used both economic indicators and text data to predict the intent of Federal Reserve

- Features:
  - Economic indicators
  - Policy Rules[1]
  - Federal Reserve published text-data



[1] Board of Governors of the Federal Reserve System: Policy Rules and How Policymakers Use Them. Link.

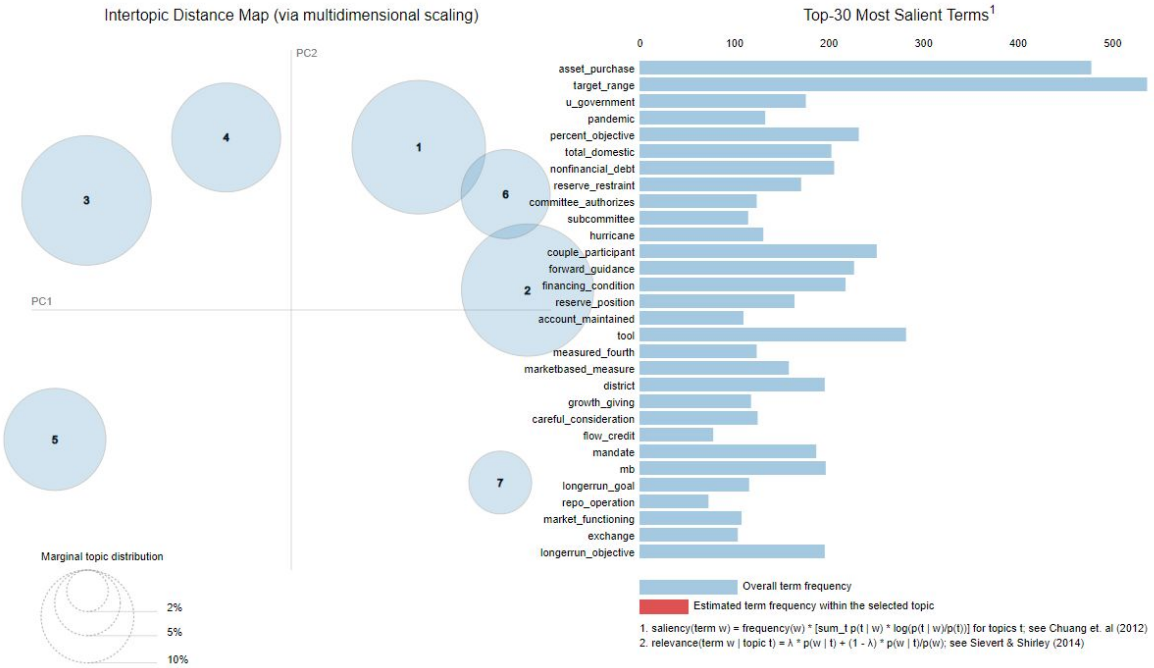
LDA Features

Data from:

- FOMC Meeting Minutes
- Fed Chairman Speeches
- FOMC Statement
- Congressional Testimony

Steps:

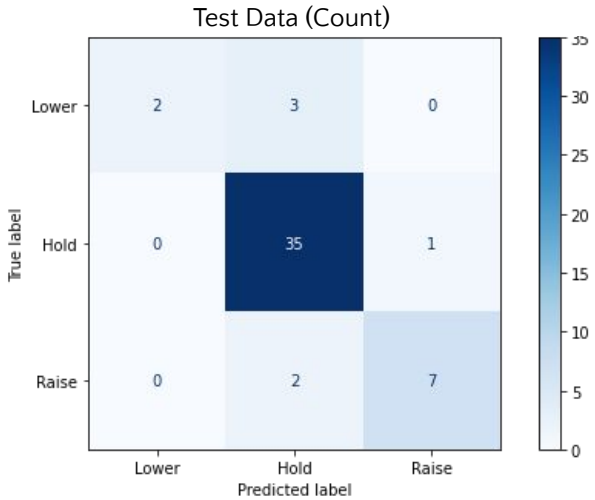
- Kept only words in policy relevant text sections
- Created bigrams -- improved topic coherence
- Chose 7 topics based on coherence score
- Topic probabilities used as features in model

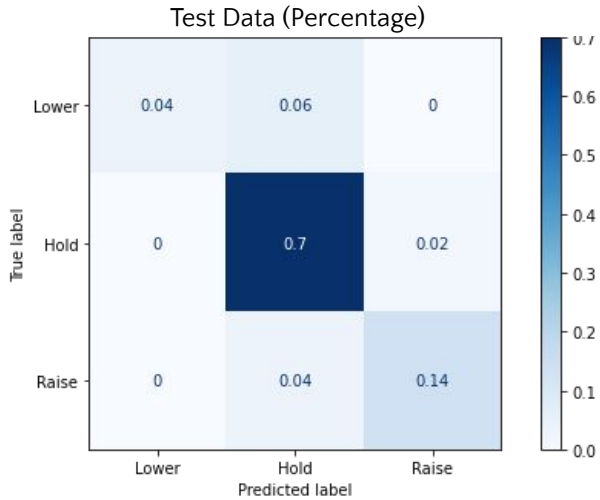


Model Prediction

- AUC Score = 0.88
- Predictions
  - Very good at predicting holding rates constant
  - Better at predicting rate increases than rate decreases

- Used as input for overall market predictions + businesses can use predictions independently



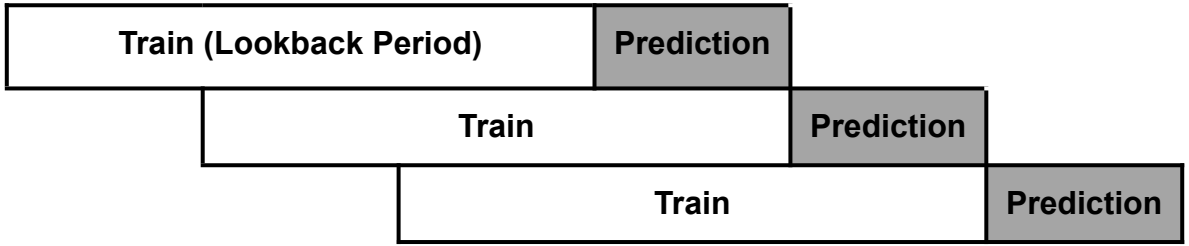


Short-Term Model: Facebook Prophet Model

Facebook Prophet is an additive model with both endogenous trend decomposition and exogenous regressors. It works best with time series that have strong seasonal effects and several seasons of historical data.

Validation Method:
- **Rolling-window Cross-Validation**

Model Improvement Methods Tried In the Process:
- **Adding Regressors (Adopted)**
- Standard Scaler
- Logistic Kernel
- Combination of Above



Feature Selection

Feature Selection:

Step 1: Use **Recursive Feature Elimination(RFE)** to identify the important features to be included as regressors in Facebook Prophet Model;

Step 2: **Eliminate multicollinearity** by dropping variables which are highly correlated.

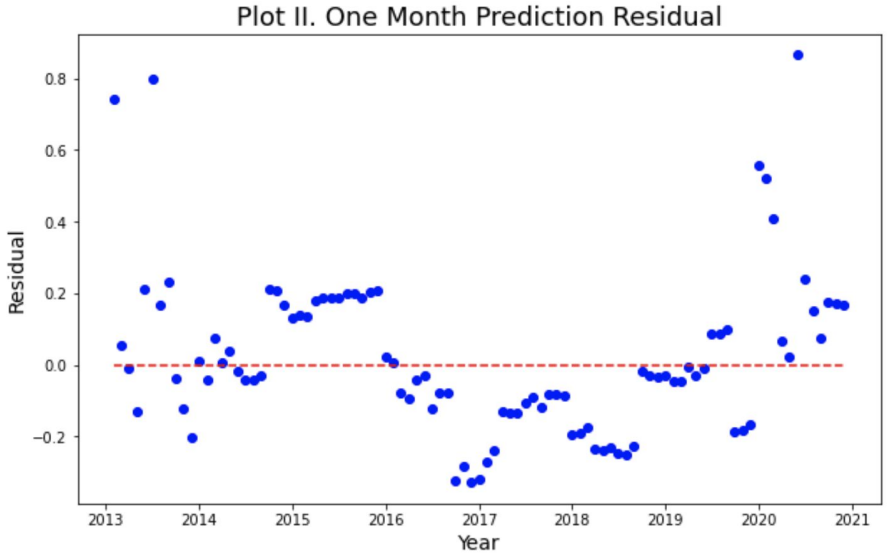
Final Independent Variables **(For further explanation, see Glossary in Appendix)****:**

- Business confidence
- CPI
- Real GDP
- Personal consumption expenditure
- Government consumption expenditure
- **Fed funds rate prediction preprocessed from Fed Minutes**
- **Recessionary probability from output of the recessionary probability model**
- **Covid Boolean**

Rolling-window Results

We have worked out 1 to 12 month rolling-window cross-validation.
The following is one-month prediction Example.
Initial(lookback period)=6 years (2190 days)
Prediction=1 month (30 days)

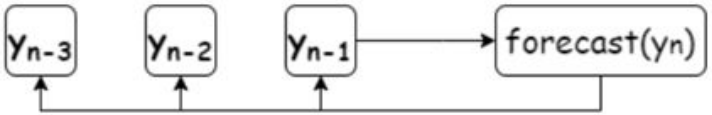
Takeaways:
- Do a nice job in capturing the trend in near future
- Suffer from previous turbulence in the lookback period



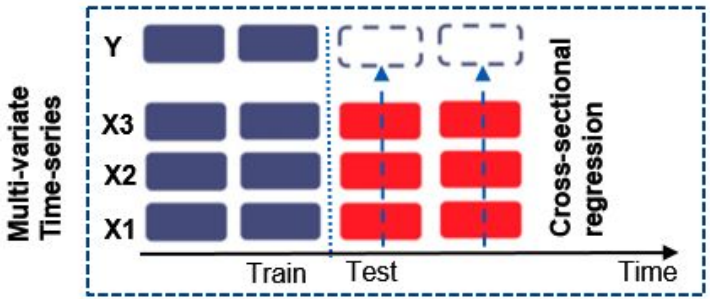
Long-Term Model

- Time-series challenge: Exogenous time-series variables do not extend to the test data set.
- Long-Term Model Steps

  Step 1. Forecast
  - One-step ahead forecasts: Auto regression
  - Multi-step ahead forecasts: Algorithm: $y_{t+h} = f_h(y_{t+h-1}, y_{t+h-2}, \cdots . y_t)$, h: forecast steps



  Step 2. Multivariate regression
  - Cross-sectional regression

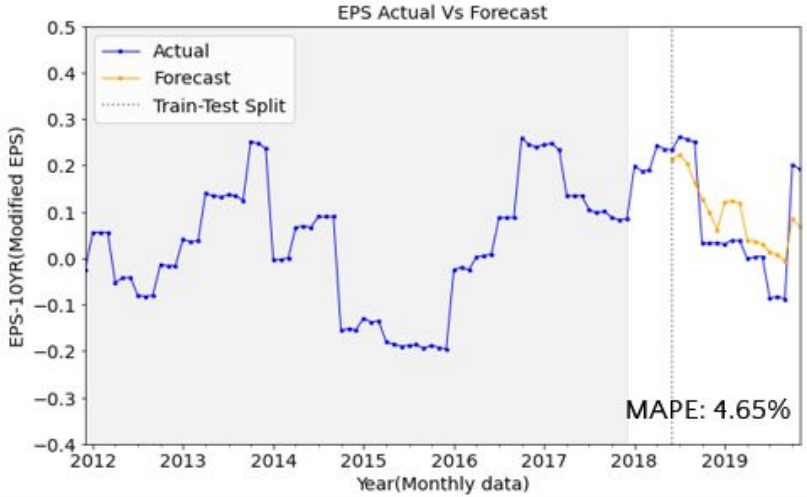


Long-Term Model

Step 3. Ensemble

- Combine Prediction Results: 1. Univariate results (Auto regression + Recursive Algorithm): U
  2. Multivariate regression: M
- Overall Prediction Results: $w_1 U + w_2 M$
  with weights $w_2 > w_1$ for forecast steps > 12

- EPS Results
  - Forecast horizon: 18months, window:6 years



Key take away:

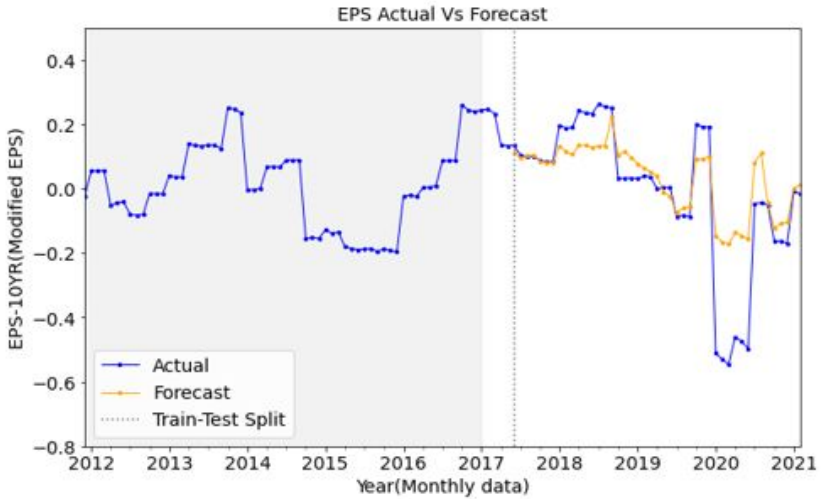
- Long-term model closely follows the target even at 18 months forecast horizon
- MAPE less than10 %

Ensemble Mechanism

- Overall Model:
  1. Short-term model: ST
  2. Long-term model: LT
  - Ensemble technique: $w_1 ST + w_2 LT$, $w_1 = 0$ for forecast steps ≥ 10
    $w_2 = 0$ for forecast steps ≤ 5

Months	1-5	6	7	8	9	10-18
w_1	1	0.6	0.5	0.4	0.3	0
w_2	0	0.4	0.5	0.6	0.7	1

- Ensemble Results



Forecast horizon: 18months,
window:6 years

S&P 500 Sector Models

Consumer Discretionary | Financials | Information Technology
Telecommunications | Healthcare

Sector-Specific Model Overview

Model

- SARIMAX -- captures seasonality and allows regressors to predict inflection points
- Features and parameters are sector-specific
  - Shifted variables used to find the highest correlation exogenous variables and time-horizon
  - Found month-over-month and year-over-year growth
  - Use overall market prediction as input

Each sector will present:

- What differentiates the sector
- Features included
- Model result + future forecast

Predictions Using Shifted X Variables

		$X2_{t-3}$	$X2_{t-2}$	$X2_{t-1}$	$X2_{t}$
	$X1_{t-3}$	$X1_{t-2}$	$X1_{t-1}$	$X1_{t}$	
Y_{t-3}	Y_{t-2}	Y_{t-1}	Y_{t}		

Sector-Specific Forecast

- Output: 18-month forecast from Sept 2021 - February 2023
- 18-month future forecasts are stitched together using the forecasts from several model horizons (example below)

	Forecast Period (Months)																	
Model Horizon	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1-Months Ahead	1																	
3-Months Ahead		2	3															
6-Months Ahead				4	5	6												
12-Months Ahead							7	8	9	10	11	12						
18-Months Ahead													13	14	15	16	17	18
Final Forecast Output	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18

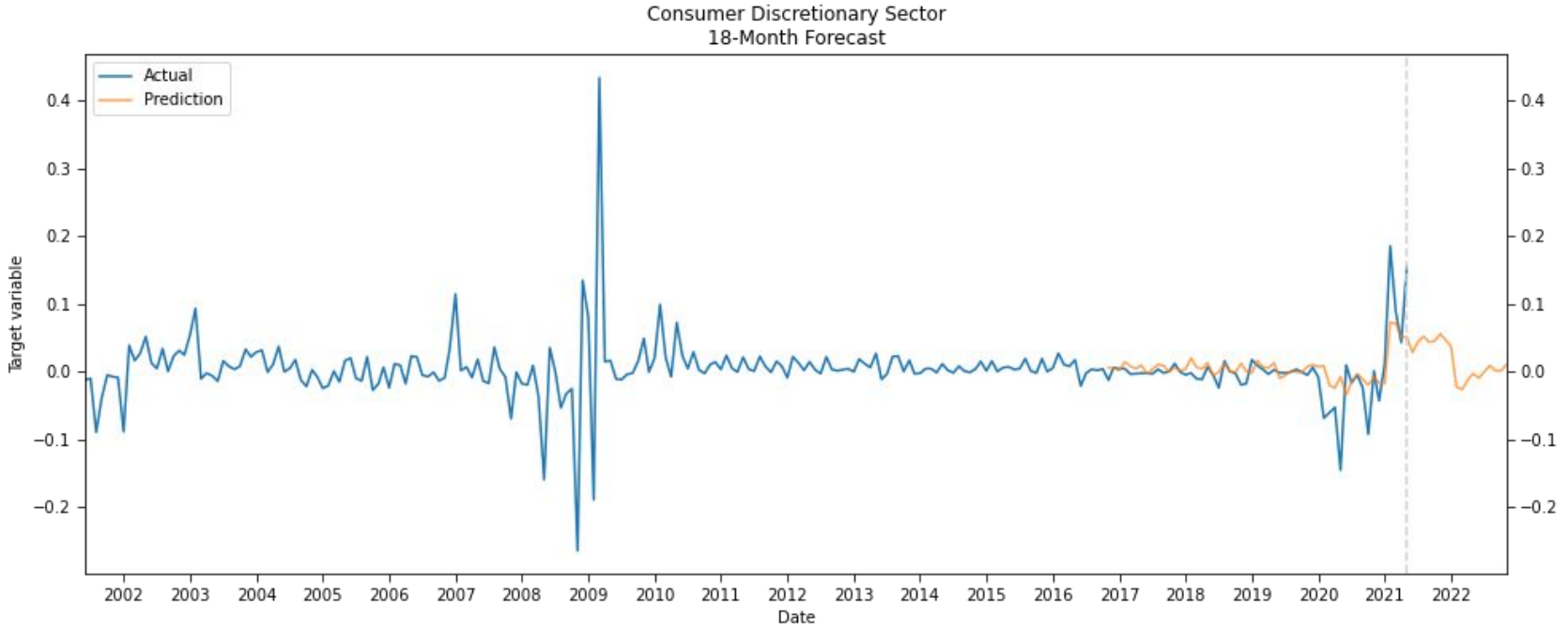
Consumer Discretionary

- Current Top Constituents (63 total): Amazon, Tesla, Home Depot, NIKE, McDonald's, Lowe's, Starbucks, Target, Booking, TJX
- Highly seasonal business that’s affected by economic downturns and consumer sentiment

Model:

- SARIMAX captures sector’s strong seasonality
- Exogenous features:
  - YoY Total Employment change
  - YoY Consumer Sentiment change

Consumer Discretionary Predictions



Metric	1-Period Ahead	12-Periods Ahead	18-Periods Ahead
MAPE	1.884	3.131	2.541
MAE	0.018	0.020	0.023

Takeaways:

- Predicts general trends. Cannot foresee severity of shocks
- Consumer sentiment is correlated with downturns
- Employment is negatively correlated with rebounds

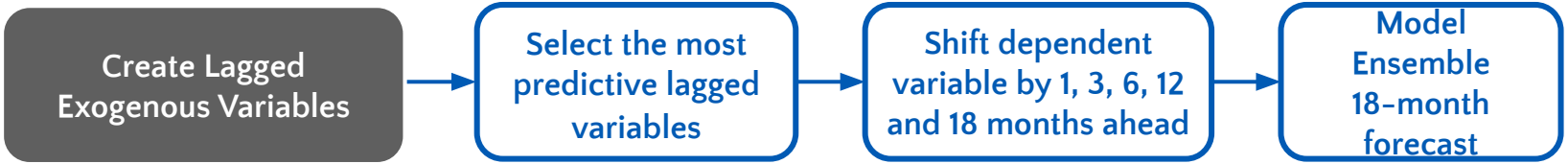
Financials Industry

SECTOR OVERVIEW

What's Special about the Financials Sector?:

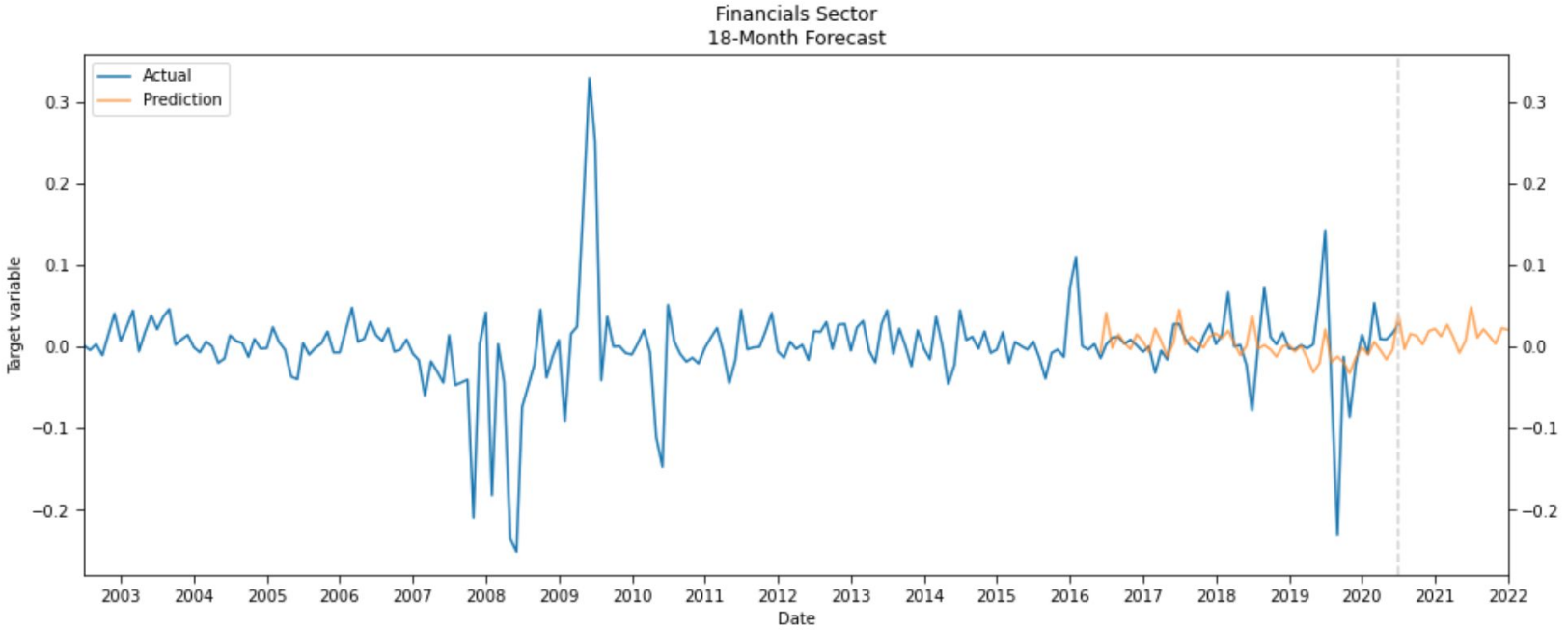
- **Banks Industry =** Largest contributor within the Financials sector
- **High domestic revenue exposure**
- **Dependent on the overall health of the US economy** since the Financial services companies' earnings portfolio is driven by the earnings of other sectors

MODEL OVERVIEW



- **Exogenous Variables:** Fed Funds Rate (FED), Business Confidence Index (BCI), Consumer Confidence Index (CCI), Unemployment rate
- **Feature Selection:** identify which variables with lags are important using correlation analysis
  - FED_11, BCI_1, PMI_1

Financials Sector – SARIMAX Time Series Model



Metrics	3-month	12-month	18-month
MAPE	2.899	2.887	3.539
MAE	0.027	0.028	0.029

Takeaways:

- Successful at capturing upward swings
- Didn't anticipate the exact magnitude of recession shock

Information Technology: Overview & Features

1. Software and Services		
Industry	Sub-Industry	Examples
Internet Software & Services	Internet Software & Services	Google, eBay, Facebook, Accenture, PayPal, Adobe, Microsoft
IT Services	IT Consulting & Other Services	
	Data Processing & Outsourced Services	
Software	Application Software	
	Systems Software	
	Home Entertainment Software	

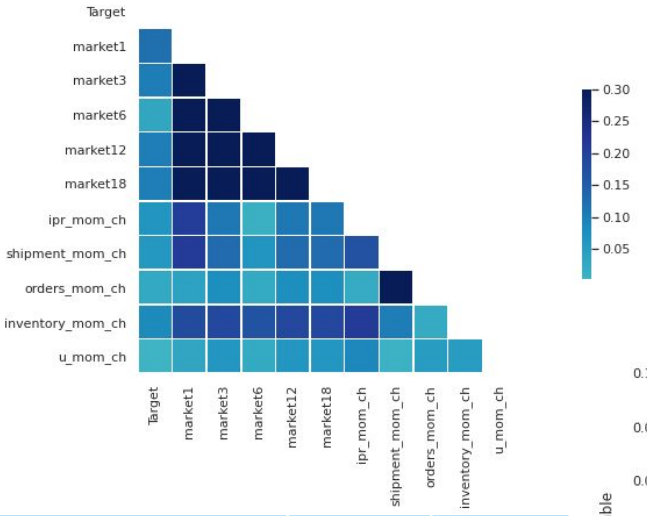
2. Technology Hardware and Equipment		
Industry	Sub-Industry	Examples
Communications Equipment	Communications Equipment	Apple, HP, Dell, Cisco Systems, SanDisk and Western Digital
Technology Hardware, Storage & Peripherals	Technology Hardware, Storage & Peripherals	
Electronic Equipment, Instruments & Components	Electronic Equipment & Instruments	
	Electronic Components	
	Electronic Manufacturing Services	
	Technology Distributions	

3. Semiconductors and Semiconductor Equipment		
Industry	Sub-Industry	Examples
Semiconductors & Semiconductor Equipment	Semiconductor Equipment	Intel, Microchip Technology, and Texas Instruments
	Semiconductors	

Potential Independent Variables

- Private fixed investment on Intellectual property products: Software (Quarterly, Quandl)
- Manufacturers' Shipments for IT Industries (Monthly, Quandl)
- Manufacturers' New Orders for IT Industries (Monthly, Quandl)
- Manufacturers' Total Inventories for IT Industries (Monthly, Quandl)
- Unemployment Rate for IT Industries (Monthly, Quandl)
- Number of Employees for IT Industries (Monthly, Quandl)

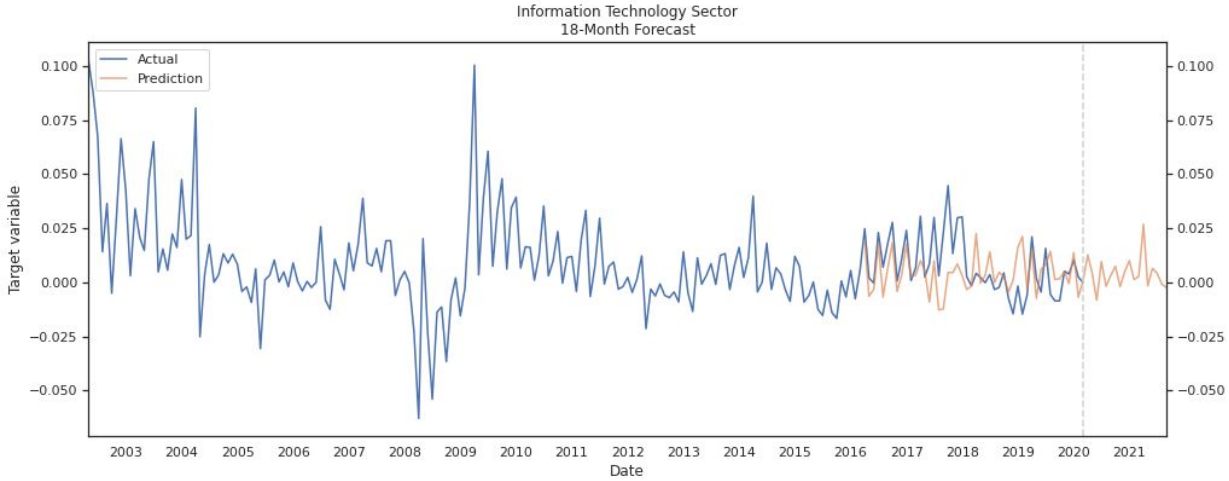
Information Technology: SARIMAX Model



Independent Variables kept:

- Private Fixed Investment on Software Intellectual Property Product,
- Manufacturers' New Orders for Information Technology Industries,
- Unemployment Rate for Information Technology Industries,
- Market Level Prediction with the respective horizon

Performance Matrix For	MAPE	MAE
1 Month Ahead	2.47567	0.01108
3 Month Ahead	5.75667	0.01078
6 Month Ahead	4.81283	0.00933
12 Month Ahead	3.52496	0.00951
18 Month Ahead	3.21647	0.00960



Telecommunications Sector

Overview

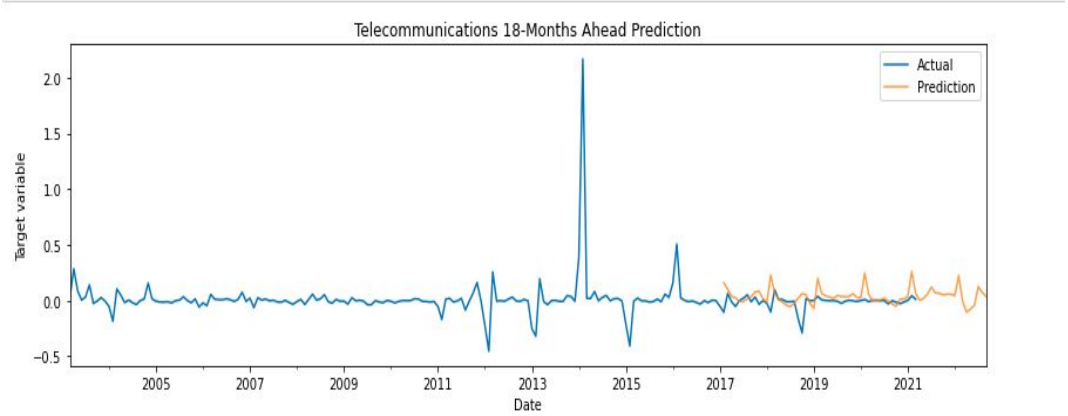
- Work style change to remote work and work from home can increase opportunities for telecommunications industry
- New Generation networks deployments could drive device upgrades and improving customer experiences

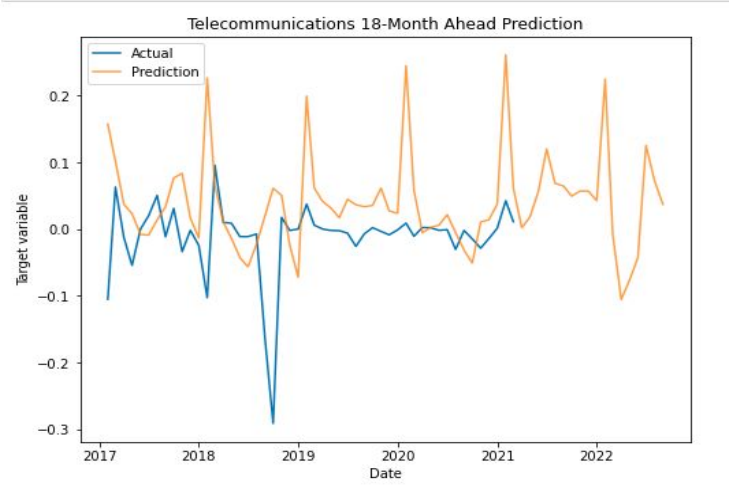
Constituents:
Verizon Communications, AT&T, T-Mobile US, etc.

Sectors	Exogenous Variables
Employment	Full-time Employee: Telecommunications
	Telecommunications Payroll
Price	Producer Price Index by Industry: Telecommunications
	Producer Price Index by Commodity: Metals and Metal Products: Copper Wire and Cable
	Consumer Price Index
International Trade	U.S. Imports of Services: Telecommunications, Computer, and Information Services
	U.S. Exports of Services: Telecommunications, Computer, and Information Services

Telecommunications Sector – Model

Metrics	1-month	6-month	12-month	18-month
MAPE	1.57	2.73	2.27	1.80
MAE	0.10	0.10	0.04	0.10





Healthcare Sector – Overview

4 Sub-Industries*

1

Services and Facilities

- Hospitals
- Facilities including nursing and residential care

Metrics & Source

1. **% aging population**
   - World Bank
2. **US national health expenditure as % of GDP**
   - Bureau of Labor Statistics

2

Medical Devices & Equipment

- Scanning machines
- Laser enabled technique apparatus

Metrics & Source

1. **# of hospitals & other medical facilities**
   - Statista
2. **AI/IOT Investment**
   - Bloomberg

3

Medical Insurance & Managed Care

- Healthcare Insurers
- Healthcare SaaS companies

Metrics & Source

1. **# of medical policyholders**
   - US Census Bureau
2. **# of medical SaaS start-ups**
   - Crunchbase

4

Pharmaceuticals

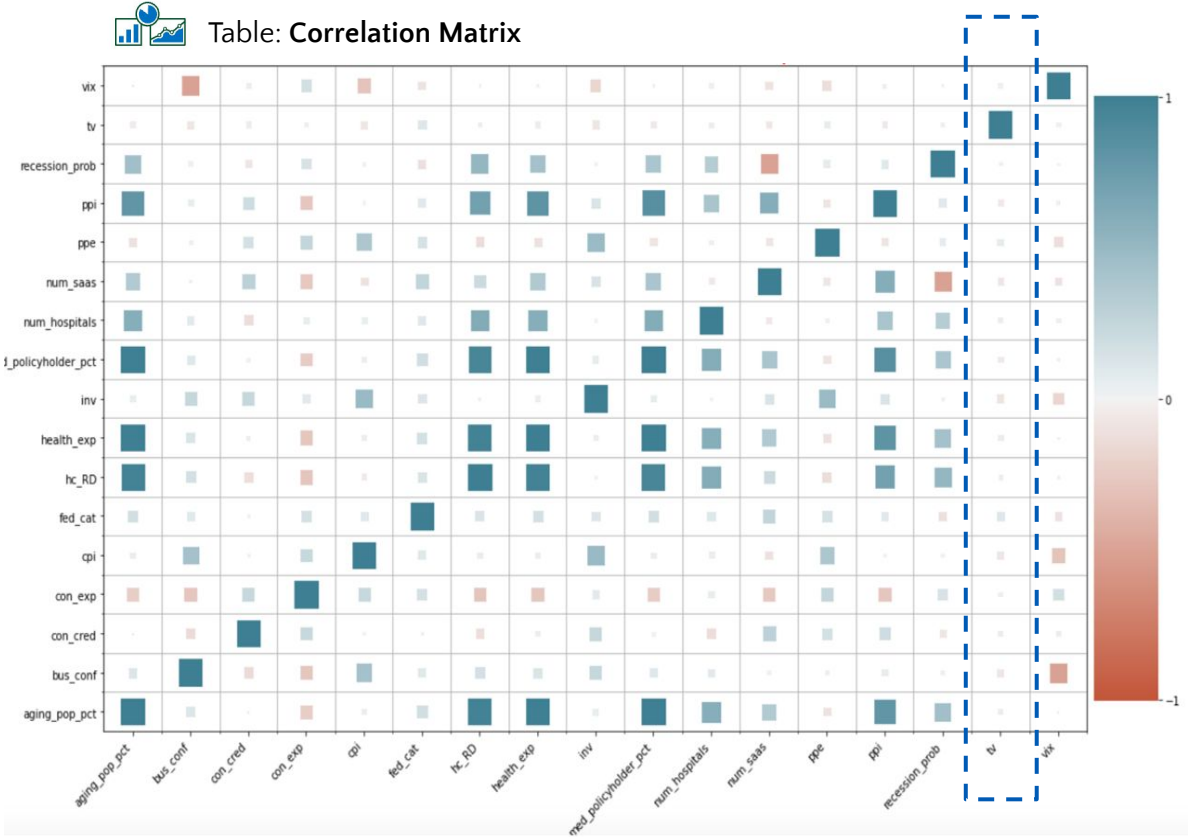
- Developing, manufacturing, and marketing drug, vaccines, etc
- Regulated by FDA

Metrics and Source

1. **Industry-wide Research & Development expenditure**
   - Congressional Budget Office
2. **Clinical trial phase**
   - FDA

Healthcare – Correlation Analysis

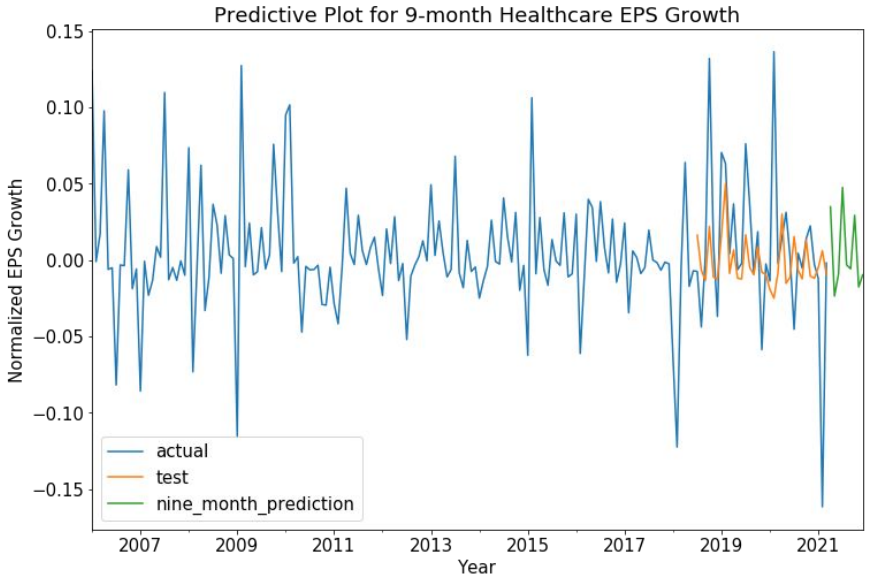
Table: **Correlation Matrix**



Predictor Variable	Target Variable
fed_cat	**0.66**
cpi	**0.57**
num_saas	**0.49**
med_policyholder_pct	**0.22**
health_exp	**0.20**
aging_pop_pct	**0.19**
hc_RD	**0.12**

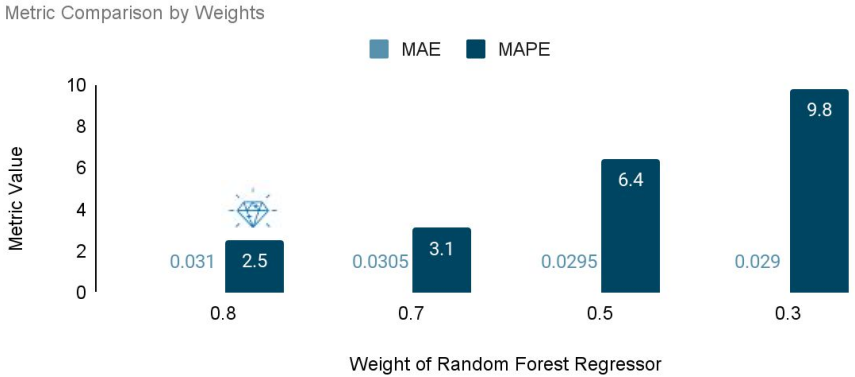
Healthcare - Modelling

Short-term: Univariate Sarimax



- MAE = 0.03
- MAPE = 1.49

Long-term: Sarimax + Random Forest Regressor



- 18-month Sarimax: -0.0073
- 18-month RF Regressor: 0.0309
- 18-month EPS growth prediction for healthcare sector would be 0.0309*0.8-0.0073*0.2 = **2.33%**

Limitations and Next Steps

Limitations + Next Steps

- Target variable -- monthly index price vs trailing P/E
- Limited leading variables for longer look-ahead periods
- Sector specific models only have data from 2001

Q&A

Appendix

Short-Term Model: Glossary for Independent Variables

- **Business confidence**: The business confidence indicator provides information on future developments, based upon opinion surveys on developments in production, orders and stocks of finished goods in the industry sector. It can be used to monitor output growth and to anticipate turning points in economic activity. Numbers above 100 suggest an increased confidence in near future business performance, and numbers below 100 indicate pessimism towards future performance. [Data Source: OECD]
- **CPI**: Consumer Price Index is a measure of change in the price level of market basket of consumer goods and services purchased. [Data Source: BLS]
- **Real GDP**: Real gross domestic product is an inflation-adjusted measure that reflects the value of all goods and services produced by an economy in a given year (expressed in base-year prices) [Data Source: Fred]
- **Personal consumption expenditure**: Personal consumption expenditure refers to a measure of imputed household expenditures defined for a period of time. Personal consumption expenditures support the reporting of the PCE Price Index, which measures price changes in consumer goods and services exchanged in the U.S. economy. [Data Source: Fred]
- **Government consumption expenditure**: Government final consumption expenditure is an aggregate transaction amount on a country's national income accounts representing government expenditure on goods and services that are used for the direct satisfaction of individual needs or collective needs of members of the community. [Data Source: Fred]
- **Covid Boolean**: We identify that Covid starts from March, 2020, therefore starting from March of 2020 the variable value equals 1, whereas previous values equal to 0.

Consumer Discretionary: Model Horizon Used for Forecast

			Forecast Period															
Model Horizon	**MAPE**	**MAE**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1-Period Ahead	1.884	0.018																		
12-Periods Ahead	3.131	0.020																		
18-Periods Ahead	2.541	0.023