

# Investing in the Future

Using market indicators to predict prices

Sawyer Shader-Seave

6.10.21

# Case Study

As an analyst for a fund:

- Find what indicators will best predict future market prices for the S&P 500 and try to predict said prices.

# Data

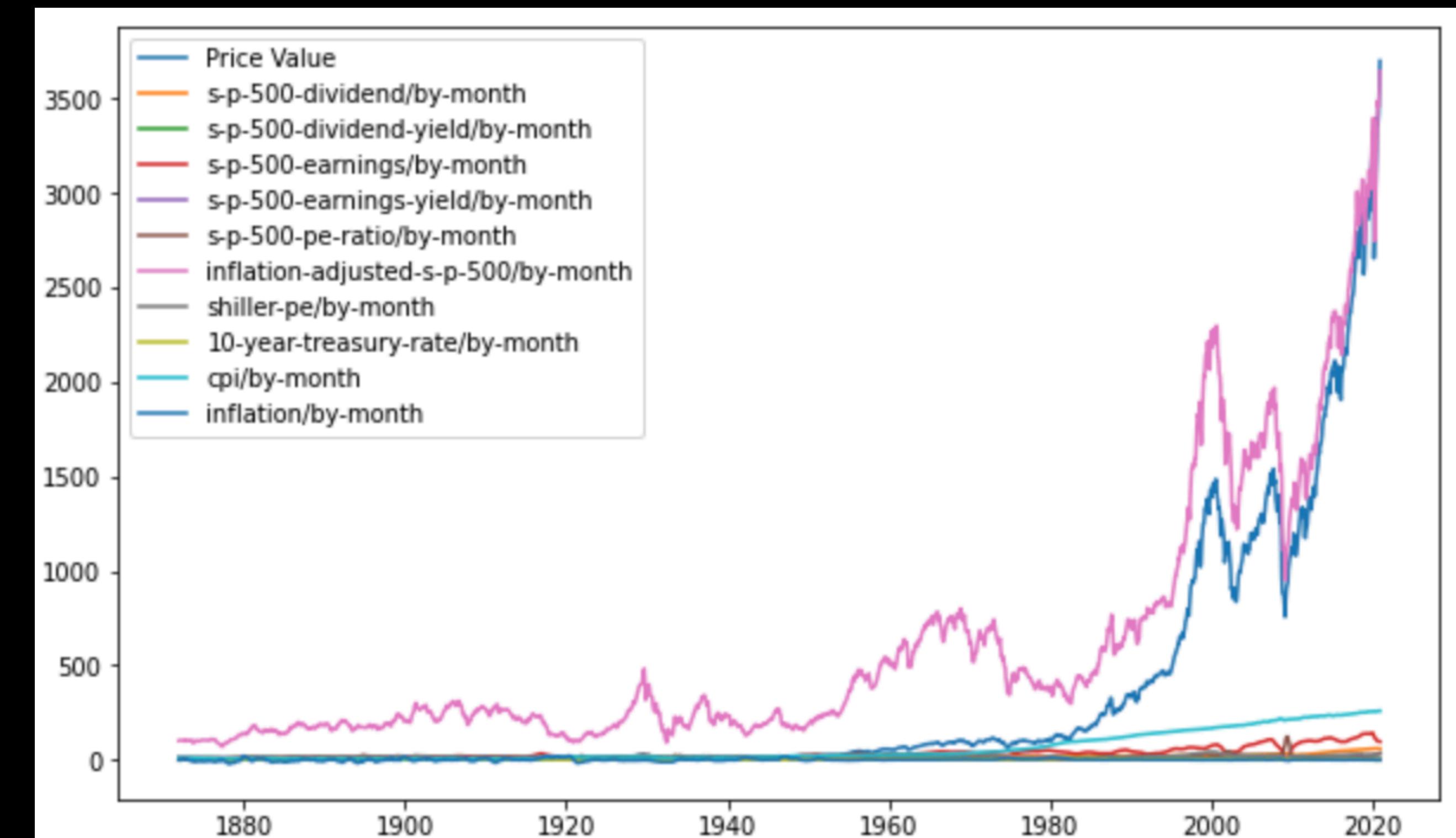
- Data
  - Source: <https://www.multpl.com/s-p-500-earnings>
  - Monthly price of the S & P 500
  - 10 economic and market variables
  - From 1872 - 2020

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 1787 entries, 2020-11-01 to 1872-01-01
Data columns (total 11 columns):
 #   Column           Non-Null Count   Dtype  
--- 
 0   Price Value      1787 non-null    float64
 1   s-p-500-dividend/by-month  1787 non-null    float64
 2   s-p-500-dividend-yield/by-month 1787 non-null    float64
 3   s-p-500-earnings/by-month     1787 non-null    float64
 4   s-p-500-earnings-yield/by-month 1787 non-null    float64
 5   s-p-500-pe-ratio/by-month    1787 non-null    float64
 6   inflation-adjusted-s-p-500/by-month 1787 non-null    float64
 7   shiller-pe/by-month       1787 non-null    float64
 8   10-year-treasury-rate/by-month 1787 non-null    float64
 9   cpi/by-month             1787 non-null    float64
 10  inflation/by-month      1787 non-null    float64
dtypes: float64(11)
memory usage: 167.5 KB
```

# EDA

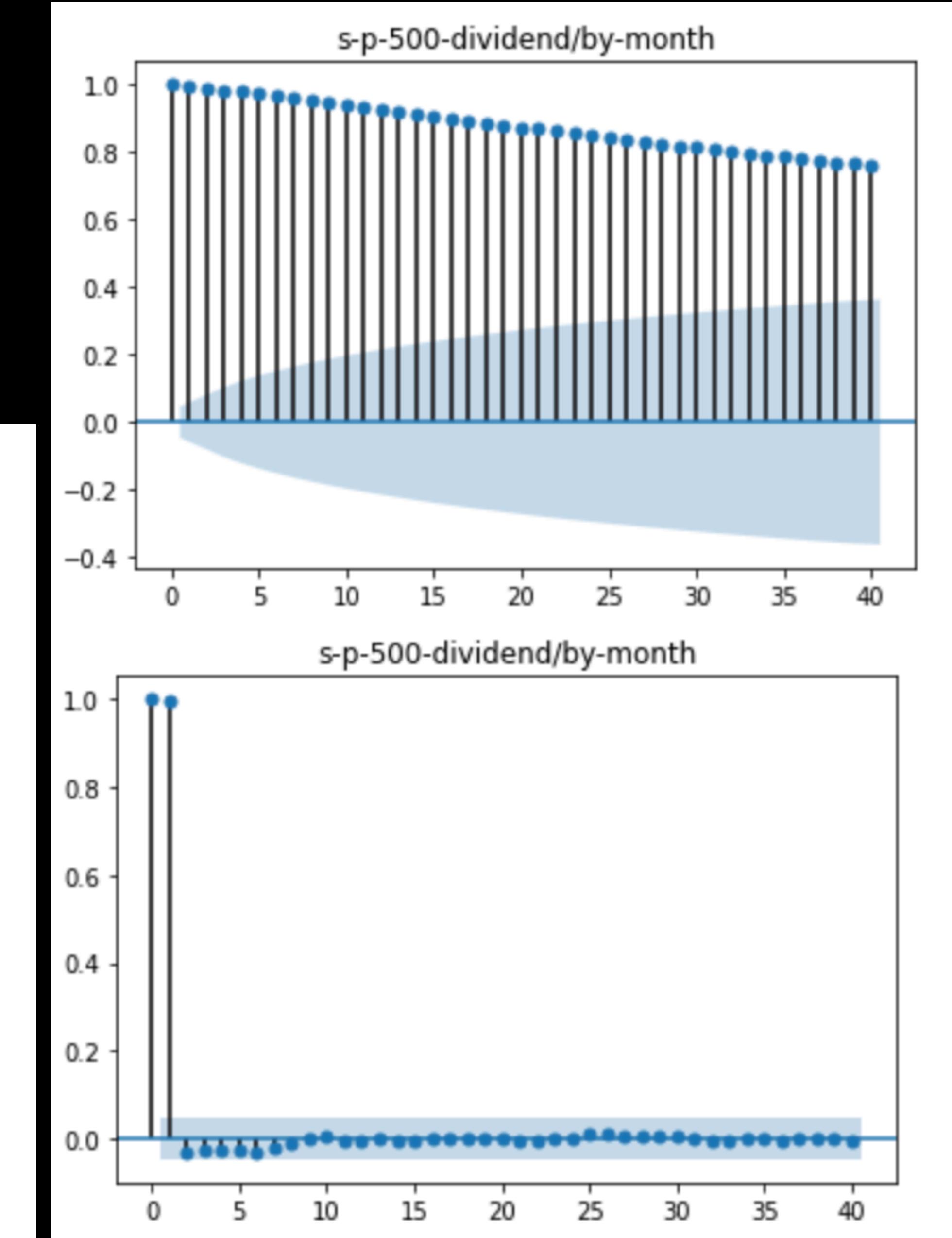
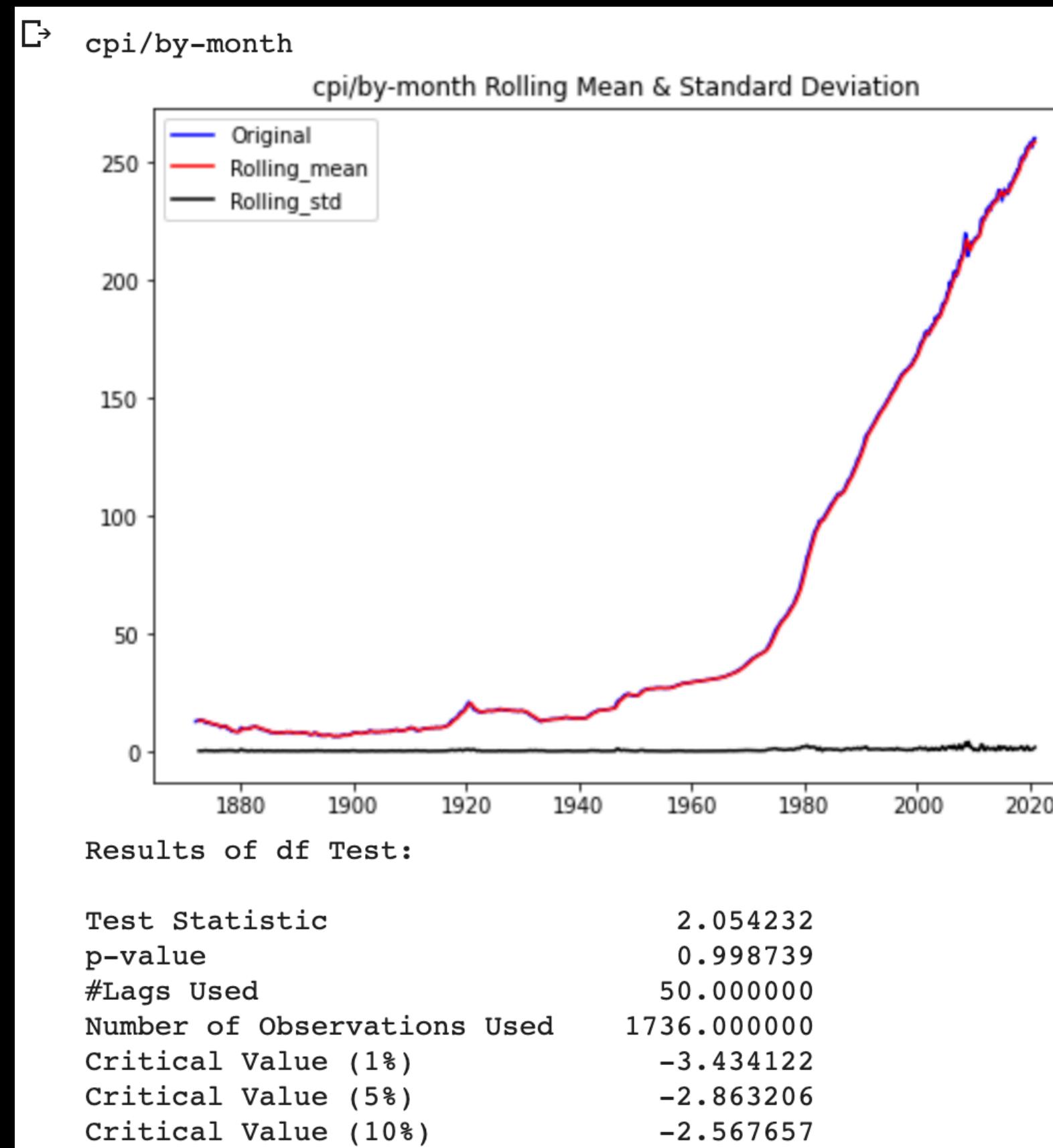
- Price is most correlated with:
  - Dividend
  - Earnings
  - Inflation
  - All but 4 of the values are not stationary

Price Value	1.000000
s-p-500-dividend/by-month	0.915552
s-p-500-dividend-yield/by-month	-0.621319
s-p-500-earnings/by-month	0.900433
s-p-500-earnings-yield/by-month	-0.437378
s-p-500-pe-ratio/by-month	0.427866
inflation-adjusted-s-p-500/by-month	0.971655
shiller-pe/by-month	0.674631
10-year-treasury-rate/by-month	-0.132116
cpi/by-month	0.904272
inflation/by-month	0.009979



# EDA autocorrelation and stationarity

- Most of the values have ACFs that stay significant for many lags
- Have PACFs drop off at 2

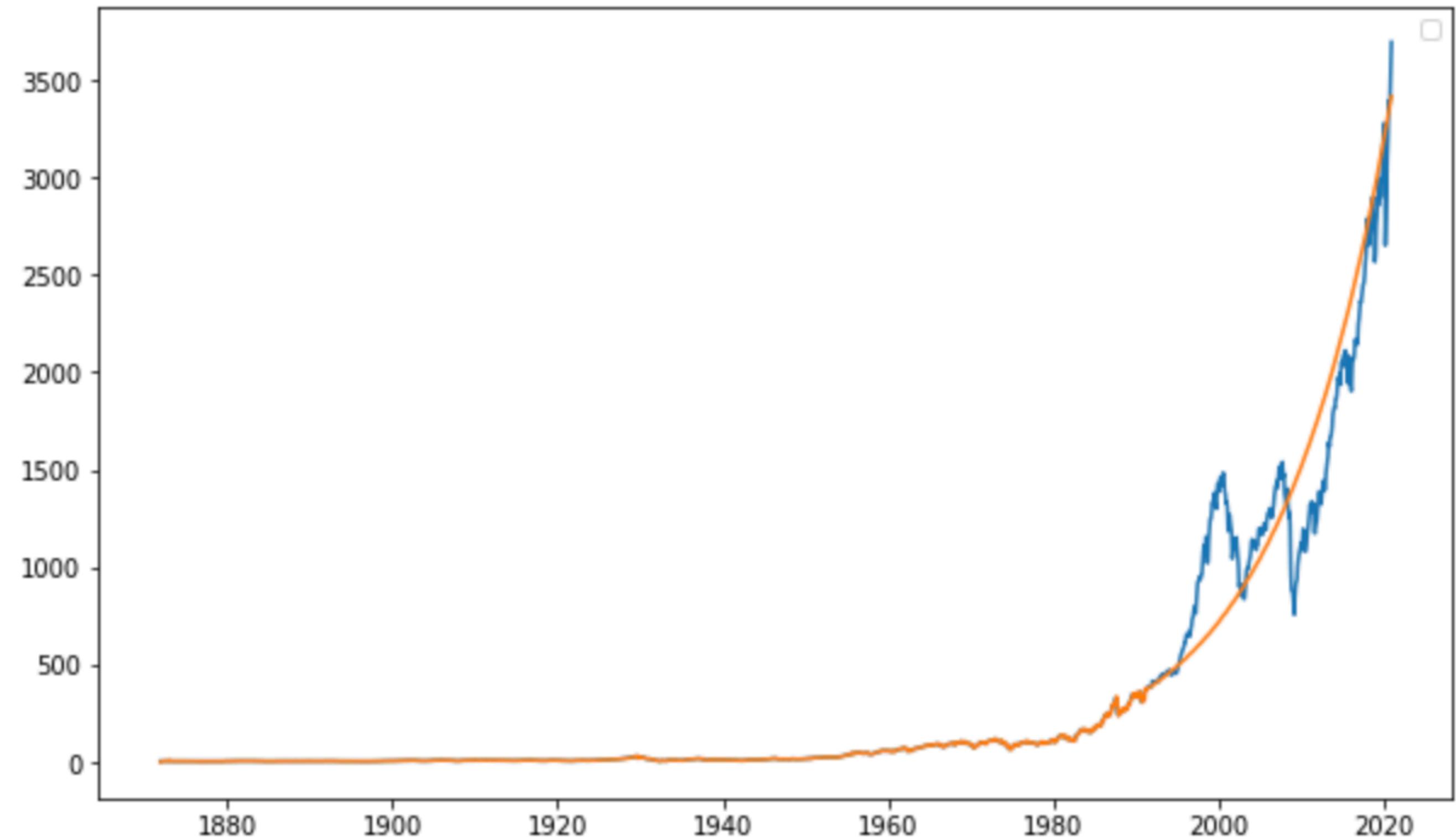


# Model

## Initial model

- ARIMA (auto)
- Initial pdq:  $(1,0,1)$  ,  $(0,1,1,12)$
- AIC: 7034.7
- Test RMSE: 311.01

```
Price Value
aic= 7034.6674  train RMSE= 2.9058  test RMSE= 311.0111
p d q
(1, 0, 1)
seasonal p d q
(0, 1, 1, 12)
```



# Model

## SARIMAX

- Baseline did the best overall
- 10 yr treasury rate
- Dividend
- All models are over fitting

	AIC	train_RMSE	test_RMSE
<b>baseline</b>	7034.67	2.9058	311.011
<b>s-p-500-dividend/by-month</b>	7111.42	2.9266	1078.95
<b>s-p-500-dividend-yield/by-month</b>	7111.5	2.9363	1083.58
<b>s-p-500-earnings/by-month</b>	7104.79	3.0493	1105.64
<b>s-p-500-earnings-yield/by-month</b>	7111.59	2.9365	1083.49
<b>s-p-500-pe-ratio/by-month</b>	7111.8	2.9438	1082.13
<b>inflation-adjusted-s-p-500/by-month</b>	7111.41	2.9616	1091.02
<b>shiller-pe/by-month</b>	7111.69	2.9654	1084.07
<b>10-year-treasury-rate/by-month</b>	7075.68	3.5063	1046.6
<b>cpi/by-month</b>	7111.16	3.0779	1070.3
<b>inflation/by-month</b>	7111.78	2.9532	1082.6

# Model

## auto-SARIMAX (multivariate)

- Variables
  - 10 yr treasury rate
  - Dividend & earnings
- AIC improved by ~0.5%
- Test RMSE improve by 5-7%

	AIC	train_RMSE	test_RMSE	pdq	Spdq
['10-year-treasury-rate/by-month', 's-p-500-dividend/by-month']	7003.31	2.9726	289.355	(1, 0, 1)	(0, 1, 1, 12)
['10-year-treasury-rate/by-month']	7001.53	2.9889	289.485	(1, 0, 1)	(0, 1, 1, 12)
['s-p-500-earnings/by-month']	7029.65	2.9182	300.207	(1, 0, 1)	(0, 1, 1, 12)
baseline	7034.67	2.9058	311.011	NaN	NaN
['s-p-500-earnings/by-month', 's-p-500-dividend/by-month']	7027.73	2.8954	311.372	(1, 0, 1)	(0, 1, 1, 12)

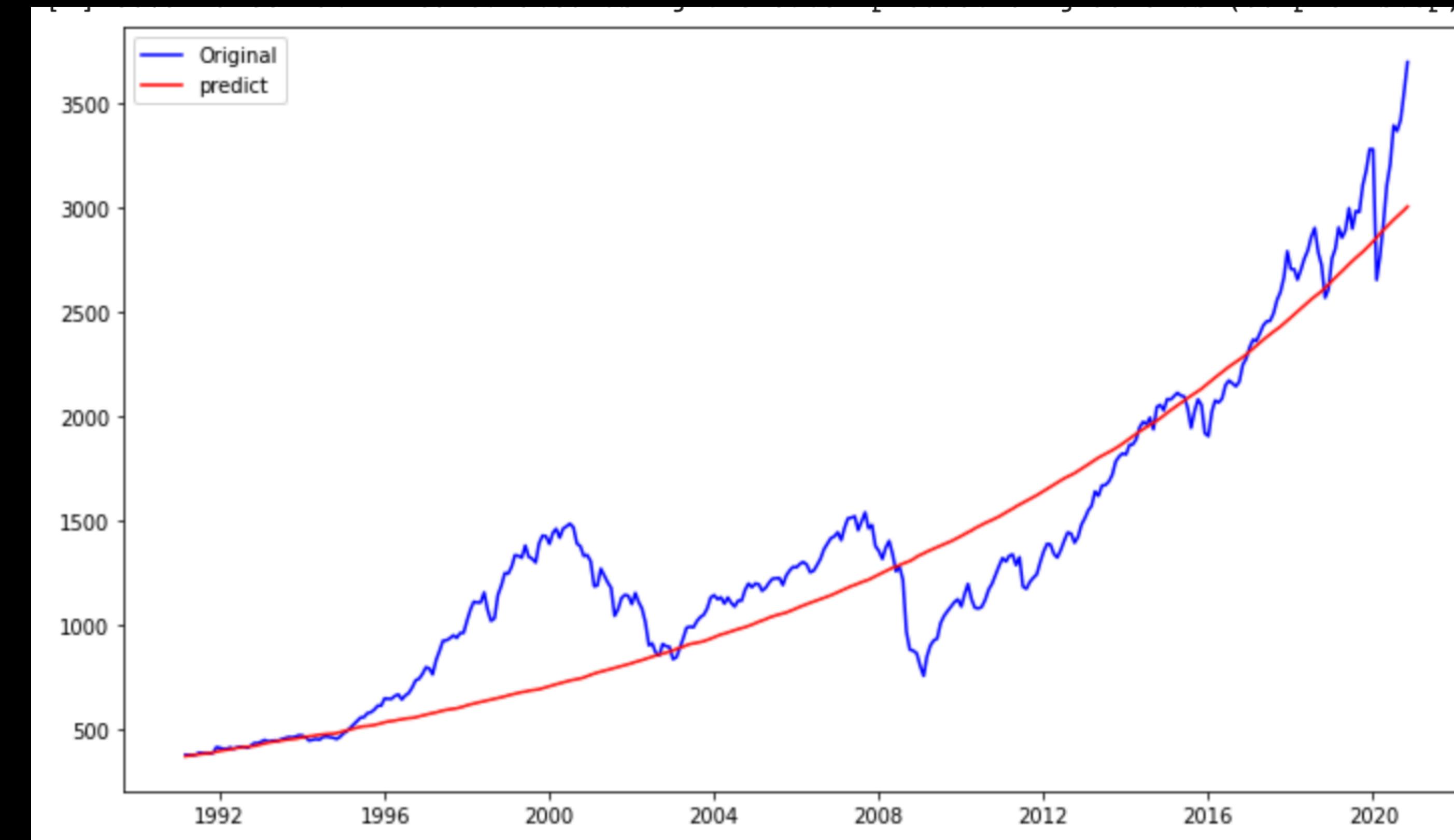
	AIC	train_RMSE	test_RMSE	pdq	Spdq
['10-year-treasury-rate/by-month', 's-p-500-earnings/by-month', 'cpi/by-month']	6990.53	3.3911	381.483	(1, 0, 1)	(0, 1, 1, 12)
['10-year-treasury-rate/by-month', 's-p-500-earnings/by-month', 's-p-500-dividend/by-month']	6996.93	2.9802	345.828	(1, 0, 1)	(0, 1, 1, 12)
['10-year-treasury-rate/by-month', 's-p-500-earnings/by-month']	6997.92	3.0309	328.436	(1, 0, 1)	(0, 1, 1, 12)
['10-year-treasury-rate/by-month', 'cpi/by-month']	6998.53	3.1699	440.346	(1, 0, 1)	(0, 1, 1, 12)
['10-year-treasury-rate/by-month', 's-p-500-earnings/by-month', 'cpi/by-month', 's-p-500-dividend/by-month']	6999.08	2.9959	1035.71	(0, 1, 1)	(0, 1, 1, 12)

# Model

## 10 Year Treasury Rate

auto-SARIMAX with 10 yr treasury rates

- 2nd best test RMSE
- 4th best AIC
- Statistically significant



Dep. Variable:	Price	Value	No. Observations:	1430		
Model:	SARIMAX(1, 0, 1)x(0, 1, 1, 12)		Log Likelihood	-3495.764		
Date:	Thu, 10 Jun 2021		AIC	7001.528		
Time:	00:00:35		BIC	7027.763		
Sample:	01-01-1872		HQIC	7011.334		
	- 02-01-1991					
Covariance Type:		opg				
	coef	std err	z	P> z		
			[0.025	0.975]		
10-year-treasury-rate/by-month	-3.0265	0.278	-10.905	0.000	-3.570	-2.483
ar.L1	1.0057	0.001	1522.627	0.000	1.004	1.007
ma.L1	0.3072	0.008	39.672	0.000	0.292	0.322
ma.S.L12	-0.9673	0.007	-146.377	0.000	-0.980	-0.954
sigma2	8.3646	0.065	129.275	0.000	8.238	8.491
Ljung-Box (L1) (Q):	0.06	Jarque-Bera (JB):	149843.20			
Prob(Q):	0.80	Prob(JB):	0.00			
Heteroskedasticity (H):	618.83	Skew:	-0.82			
Prob(H) (two-sided):	0.00	Kurtosis:	53.58			

- AIC improved by 0.47%
- Test RMSE improved by 6.9%

# Model 2

## Facebook Prophet

- RMSE more than 2X ARIMA model
- Best variable:
  - Inflation
  - CPI
  - Treasury rate



	train_RMSE	test_RMSE
inflation-adjusted-s-p-500/by-month	9.3469	976.626
cpi/by-month	19.77	1184.16
10-year-treasury-rate/by-month	14.7064	1223
s-p-500-dividend/by-month	19.7244	1257.55
shiller-pe/by-month	15.839	1268.77
s-p-500-pe-ratio/by-month	18.1629	1294.62
s-p-500-earnings-yield/by-month	17.9411	1307.69
s-p-500-dividend-yield/by-month	19.6504	1315.23
inflation/by-month	21.2222	1329.45
baseline	21.5287	1332.03
s-p-500-earnings/by-month	21.5214	1337.73

# Model 2

## Facebook Prophet combinations

- Best variables
  - Inflation
  - CPI
  - 10 yr treasury
- 10% better than prior models
- Still worse than ARIMA baseline

	train_RMSE	test_RMSE
['inflation-adjusted-s-p-500/by-month', 'cpi/by-month', '10-year-treasury-rate/by-month']	6.9302	891.681
['inflation-adjusted-s-p-500/by-month', 'cpi/by-month', '10-year-treasury-rate/by-month', 's-p-500-dividend/by-month']	6.9428	892.799
['inflation-adjusted-s-p-500/by-month', 'cpi/by-month']	9.221	899.745
['inflation-adjusted-s-p-500/by-month', 'cpi/by-month', 's-p-500-dividend/by-month']	9.2851	901.443
['cpi/by-month', '10-year-treasury-rate/by-month', 's-p-500-dividend/by-month']	10.2073	974.124
['inflation-adjusted-s-p-500/by-month']	9.3469	976.626
['inflation-adjusted-s-p-500/by-month', 's-p-500-dividend/by-month']	9.4031	978.611
['cpi/by-month', '10-year-treasury-rate/by-month']	10.5328	987.305
['inflation-adjusted-s-p-500/by-month', '10-year-treasury-rate/by-month', 's-p-500-dividend/by-month']	8.3913	1003.33
['inflation-adjusted-s-p-500/by-month', '10-year-treasury-rate/by-month']	8.4322	1005.25
['cpi/by-month', 's-p-500-dividend/by-month']	17.9469	1085.78
['10-year-treasury-rate/by-month', 's-p-500-dividend/by-month']	13.6491	1176.85
['cpi/by-month']	19.77	1184.16
['10-year-treasury-rate/by-month']	14.7064	1223
['s-p-500-dividend/by-month']	19.7244	1257.55
baseline	21.5287	1332.03

# Conclusions

- Models failed to predict future prices accurately enough
- 10 year treasury rates
  - improved all models
  - Statistically significant
  - Coefficient: -2.7 to -3.1
- Earnings, inflation, and CPI
- Variables can/do improve price prediction models:
  - best model was 33.1% better than baseline

# Future Work

- Run using different models
- Find more explanatory variables
- Focus on shorter timespans

Thank You  
Questions?