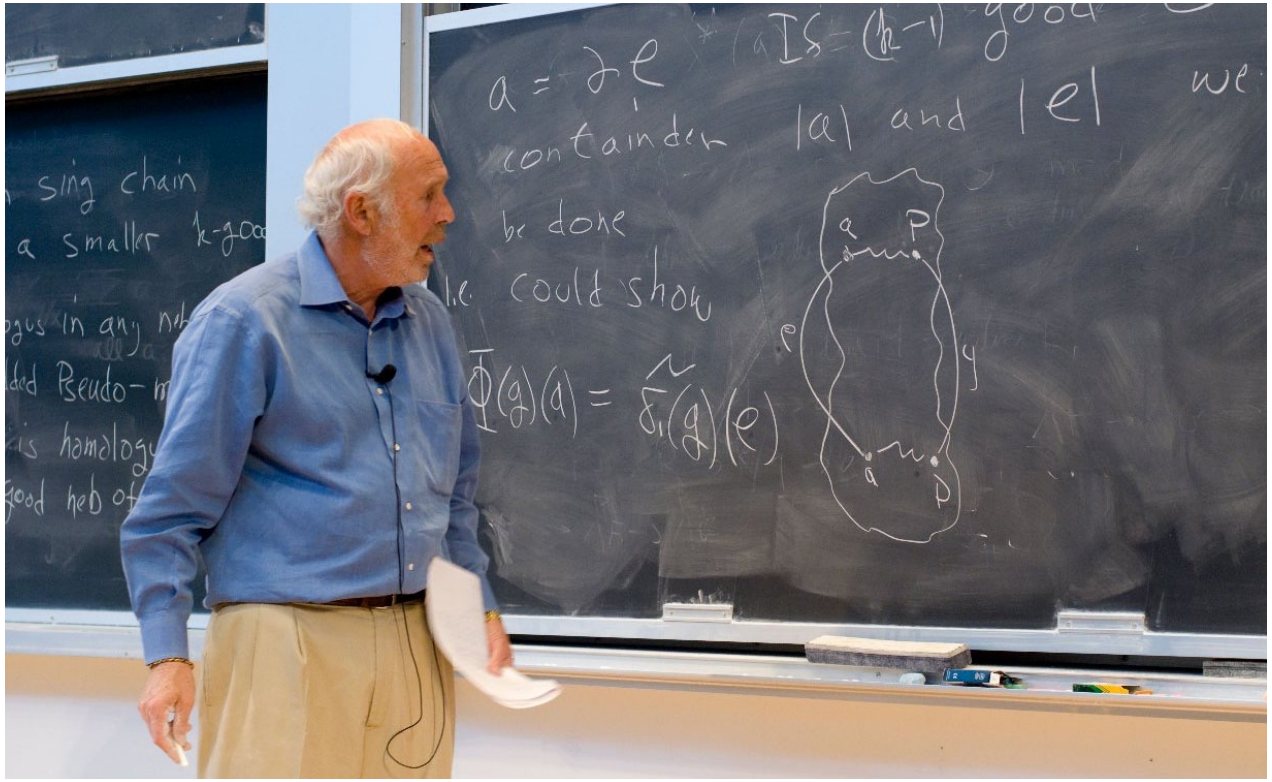
Stock Prediction
Using Autoregressive Modeling

CSC-475: Seminar in Computer Science

Emil Westling

FURMAN UNIVERSITY

Background



- Jim Simons, PhD
- 39.1% 1988-2018

Algorithmic Trading:

- 70% of intra-day trading volume

- Quantitative Researcher:
- $250,000+/yr

Problem

HOW CAN AUTOREGRESSIVE MODELS BE USED MOST EFFECTIVELY FOR STOCK PREDICTION?

Data

- Apple (7.29%) & Amazon (3.51%)
- yfinance

Date	Open	High	Low	Close	Volume	Dividends	Stock Splits
2021-09-30	141.830138	142.540968	139.480448	139.697647	89056700	0.0	0.0
2021-10-01	140.092561	141.099573	137.338105	140.833008	94639600	0.0	0.0
2021-10-04	139.954314	140.398594	136.508778	137.367691	98322000	0.0	0.0
2021-10-05	137.713246	140.428218	137.584898	139.312607	80861100	0.0	0.0
2021-10-06	137.693511	140.339368	136.607516	140.191284	83221100	0.0	0.0

Tweet Data

- eng_spacysentiment
- Sum up each day

	Date	Tweet	Stock Name	Company Name
Date				
2022-09-29 22:23:54	2022-09-29 22:23:54	$NIO just because I'm down money ...	AAPL	Apple Inc.
2022-09-29 20:37:01	2022-09-29 20:37:01	After trading for 9+ years \n\nThis is ...	AAPL	Apple Inc.
2022-09-29 20:19:43	2022-09-29 20:19:43	Not something you see very often...	AAPL	Apple Inc.
2022-09-29 20:13:48	2022-09-29 20:13:48	$AAPL was down almost 5% today...	AAPL	Apple Inc.
2022-09-29 19:50:00	2022-09-29 19:50:00	$AAPL APPLE JUST FIRED ITS HEAD...	AAPL	Apple Inc.

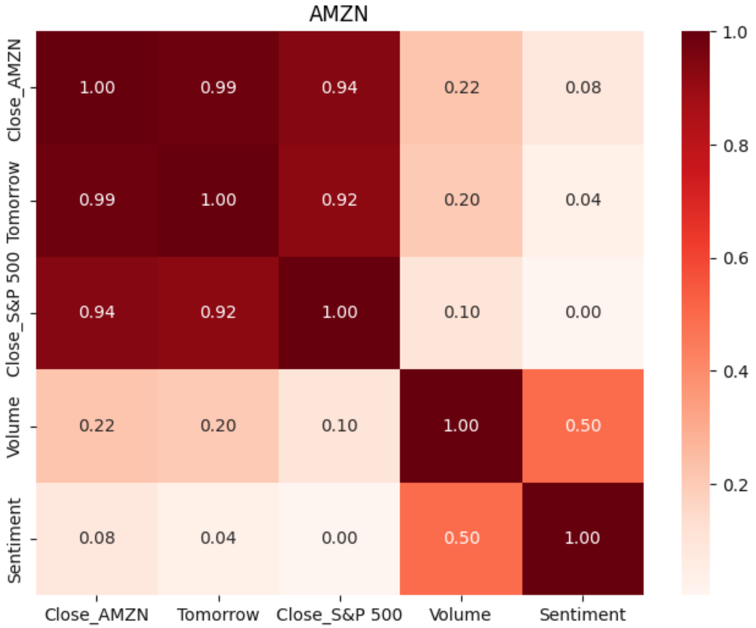
Final Data Frames

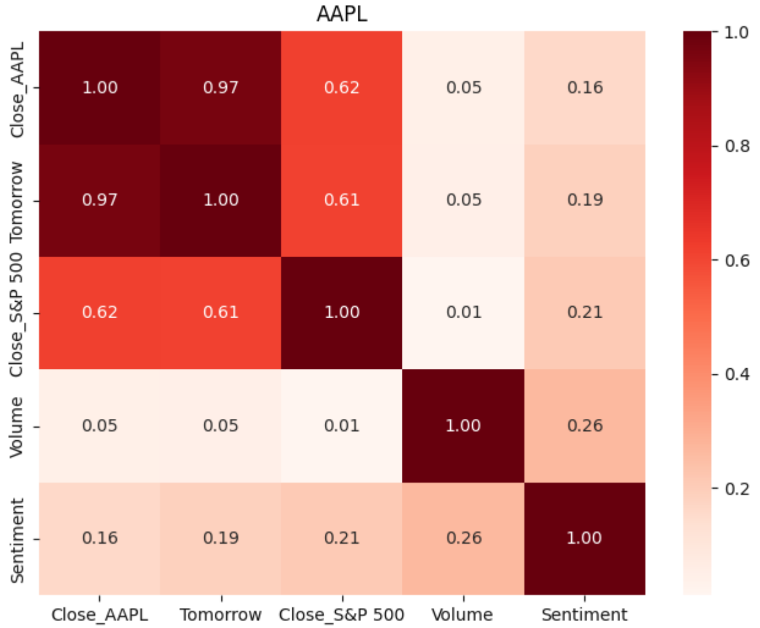
Date	Open	High	Low	Close	Volume	Tomorrow	direction	Sentiment
2021-09-30	141.830138	142.540968	139.480448	139.697647	89056700	140.833008	1	4
2021-10-01	140.092561	141.099573	137.338105	140.833008	94639600	137.367691	-1	3
2021-10-04	139.954314	140.398594	136.508778	137.367691	98322000	139.312607	1	4
2021-10-05	137.713246	140.428218	137.584898	139.312607	80861100	140.191284	1	0
2021-10-06	137.693511	140.339368	136.607516	140.191284	83221100	141.464828	1	2

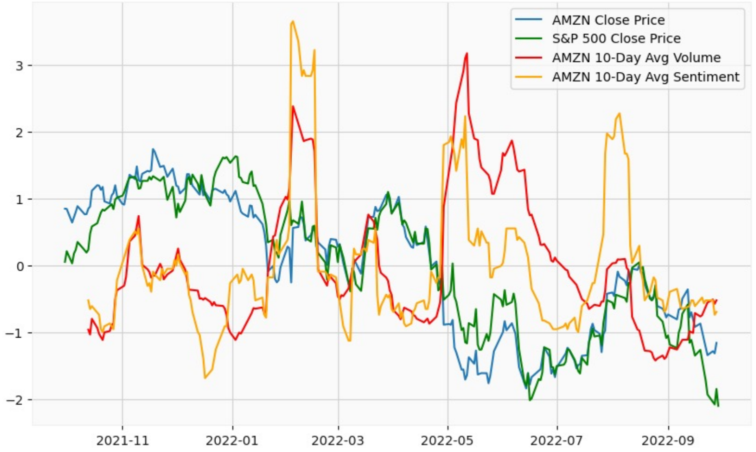
Data Visualization

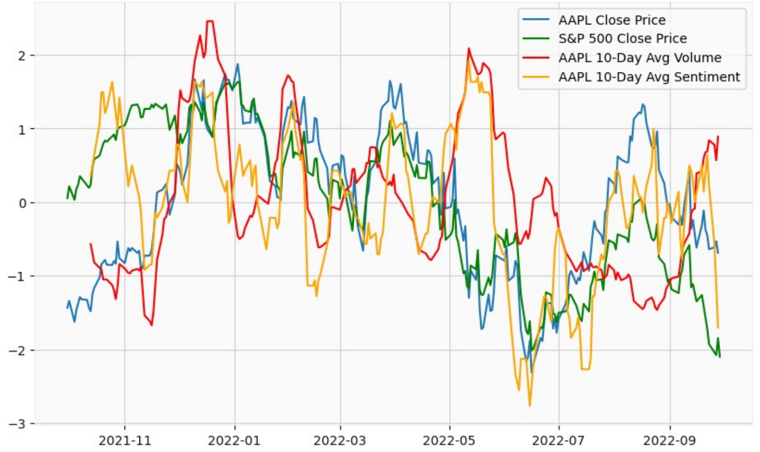
AAPL

AMZN









Correlations

Model

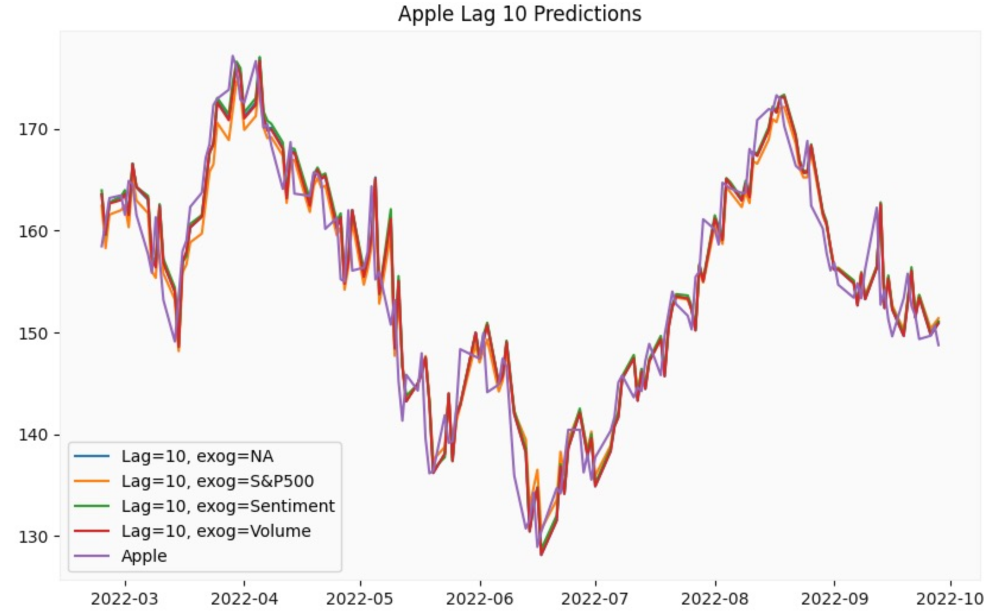
- Statsmodels' AutoReg
- Price & Direction

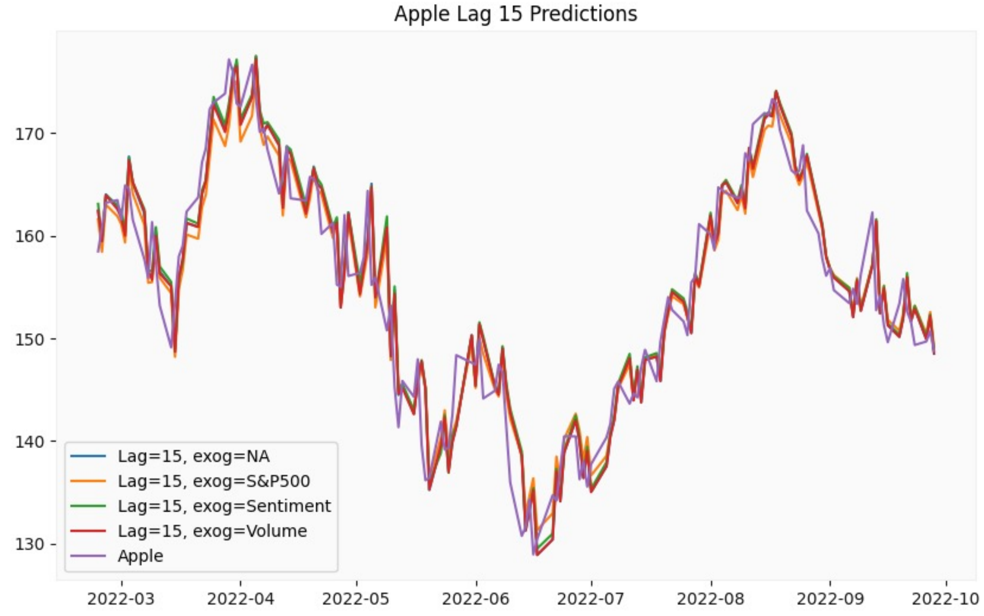
$$y_t = \delta_0 + \delta_1 t + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \sum \kappa_j x_{t,j} + \varepsilon_t$$

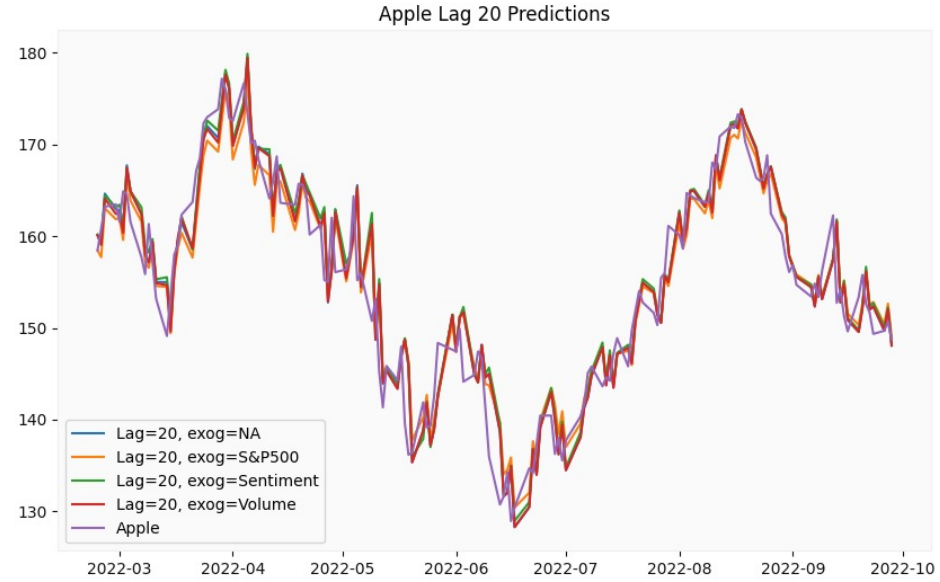
Forecasting

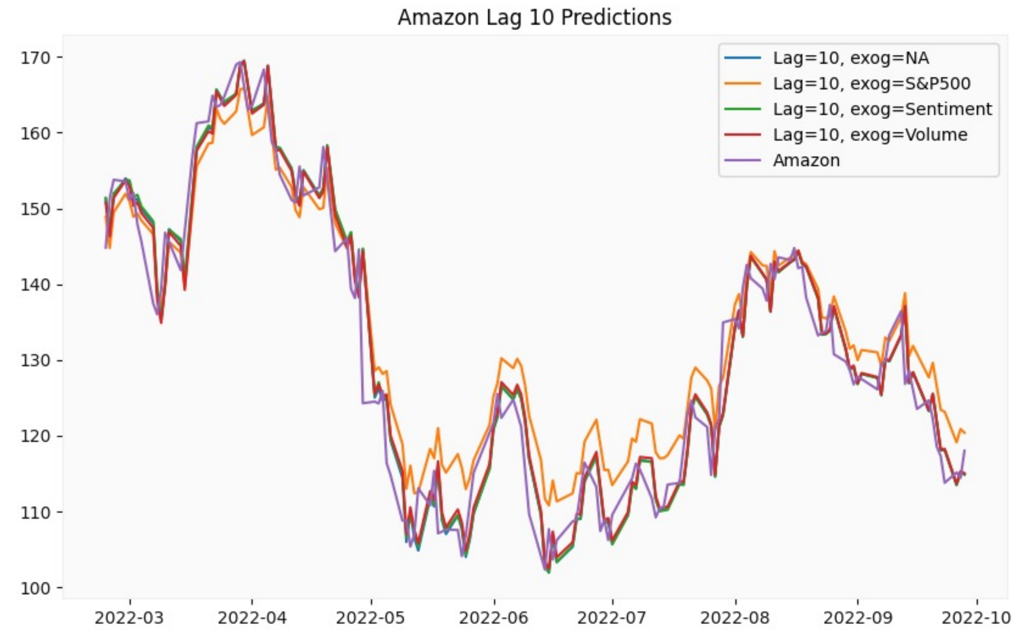
- Lags: 10, 15, 20
- Exogenous:
  - S&P 500,
  - Sentiment
  - Volume

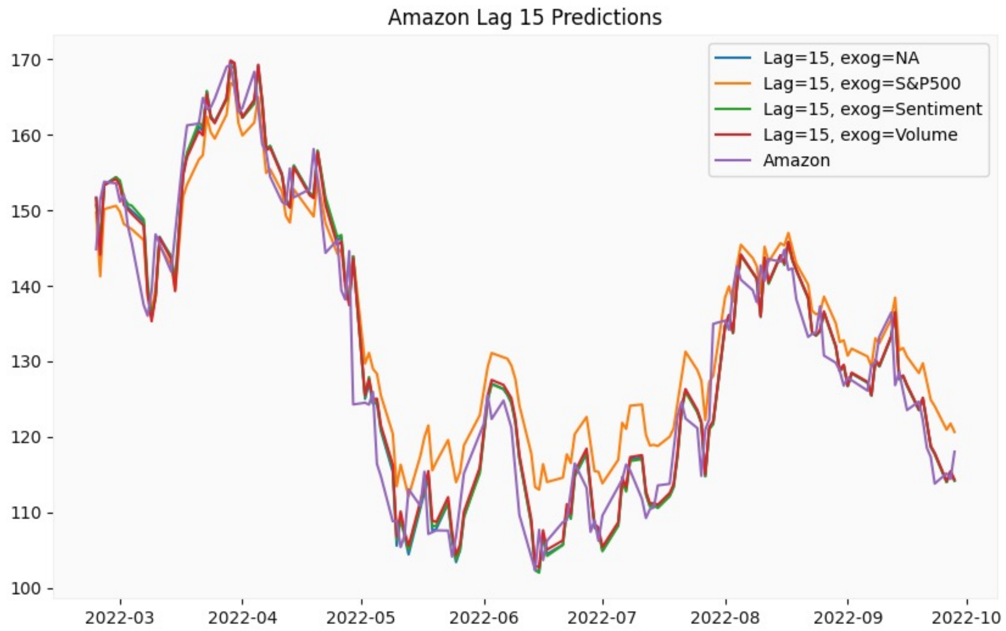
	Company	Amazon		Apple	
		MAE_Direction	MAE_Price	MAE_Direction	MAE_Price
Lag	Exog				
10	NA	1.019868	**3.346440**	0.953642	2.762461
	S&P500	**0.993377**	4.859555	**0.940397**	2.795329
	Sentiment	1.033113	3.356564	0.953642	**2.744114**
	Volume	1.006623	3.360806	0.953642	2.755890
15	NA	1.033113	3.411418	0.966887	2.839545
	S&P500	**0.993377**	5.569160	1.059603	2.787392
	Sentiment	1.006623	3.413289	1.006623	2.829494
	Volume	1.006623	3.443118	0.980132	2.835158
20	NA	1.086093	3.424944	0.993377	2.853466
	S&P500	**0.993377**	5.841501	0.980132	2.768990
	Sentiment	1.033113	3.474168	0.953642	2.837408
	Volume	1.086093	3.447584	0.980132	2.851404

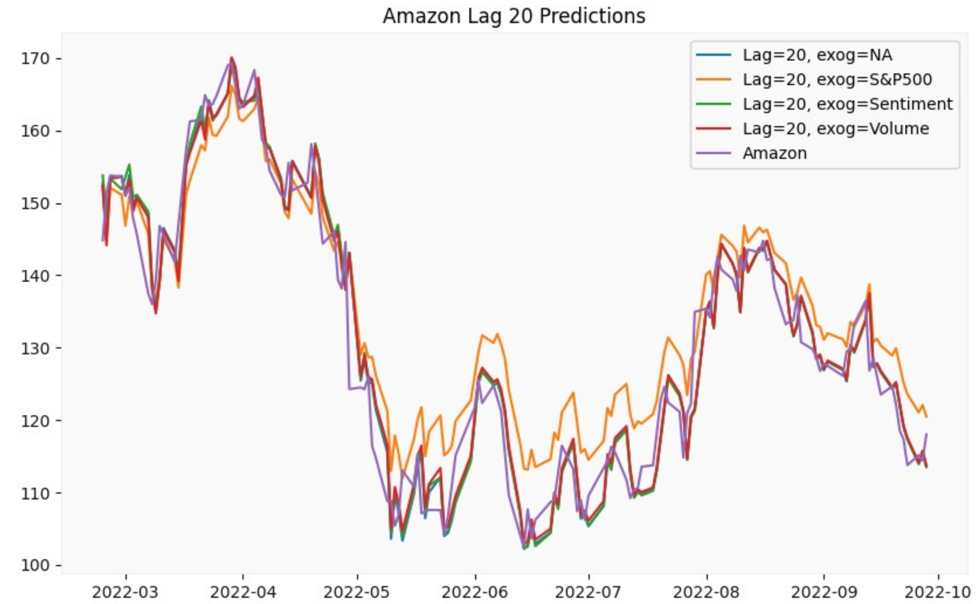

Apple Lag 10 Predictions
170
160
150
140
130
Lag=10, exog=NA
Lag=10, exog=S&P500
Lag=10, exog=Sentiment
Lag=10, exog=Volume
Apple
2022-03
2022-04
2022-05
2022-06
2022-07
2022-08
2022-09
2022-10


Apple Lag 15 Predictions
170
160
150
140
130
Lag=15, exog=NA
Lag=15, exog=S&P500
Lag=15, exog=Sentiment
Lag=15, exog=Volume
Apple
2022-03
2022-04
2022-05
2022-06
2022-07
2022-08
2022-09
2022-10


Apple Lag 20 Predictions
180
170
160
150
140
130
Lag=20, exog=NA
Lag=20, exog=S&P500
Lag=20, exog=Sentiment
Lag=20, exog=Volume
Apple
2022-03
2022-04
2022-05
2022-06
2022-07
2022-08
2022-09
2022-10


Amazon Lag 10 Predictions
Lag=10, exog=NA
Lag=10, exog=S&P500
Lag=10, exog=Sentiment
Lag=10, exog=Volume
Amazon
170
160
150
140
130
120
110
100
2022-03
2022-04
2022-05
2022-06
2022-07
2022-08
2022-09
2022-10


Amazon Lag 15 Predictions
Lag=15, exog=NA
Lag=15, exog=S&P500
Lag=15, exog=Sentiment
Lag=15, exog=Volume
Amazon
170
160
150
140
130
120
110
100
2022-03
2022-04
2022-05
2022-06
2022-07
2022-08
2022-09
2022-10


Amazon Lag 20 Predictions
Lag=20, exog=NA
Lag=20, exog=S&P500
Lag=20, exog=Sentiment
Lag=20, exog=Volume
Amazon
170
160
150
140
130
120
110
100
2022-03
2022-04
2022-05
2022-06
2022-07
2022-08
2022-09
2022-10

Conclusions

- Exogenous variables does not improve the models significantly
- Lower lag -> Better Performance
- Most predictions are similar
- ARIMA, VaR, Markov-Chains, DLR

https://emilwestling.com/AlgorithmTrading