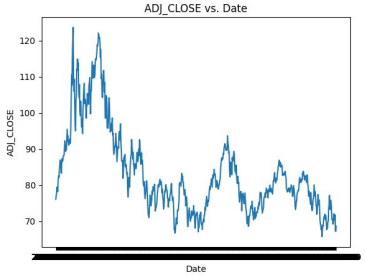
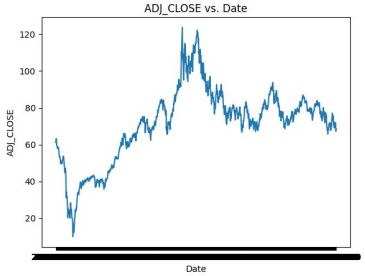
Visualizations for EDA of Time Series and Results

Julia Gugulski
CSCI 1420 Capstone Project
Spring 2025

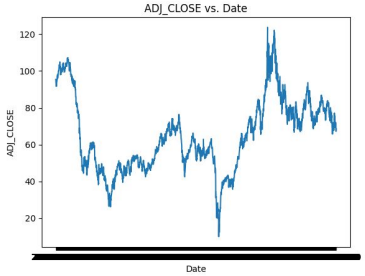
Raw data plot:



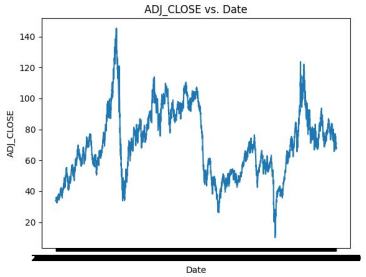
3 years



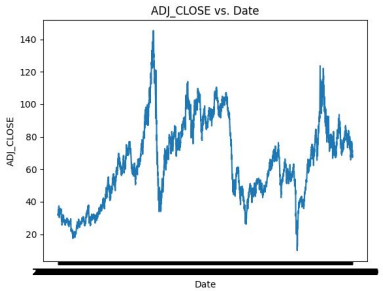
5 years



10 years



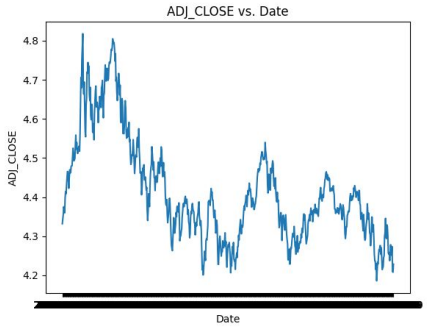
20 years



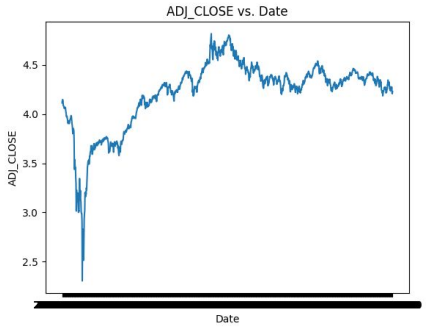
all

Transformed data plot:

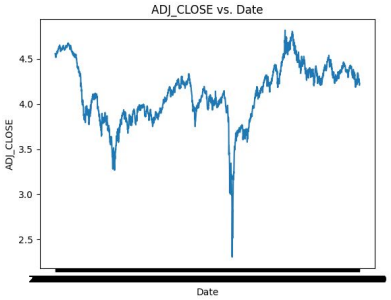
The transformation = logarithm of each adjusted closing value



3 years

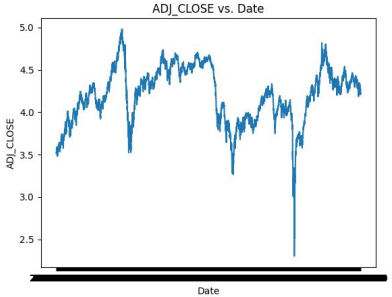


5 years

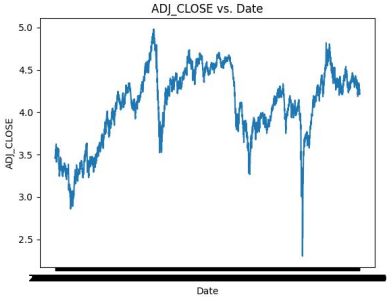


10 years

20 years

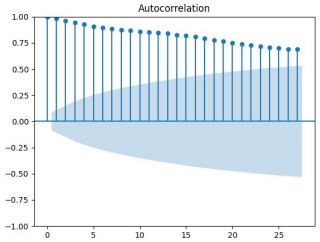


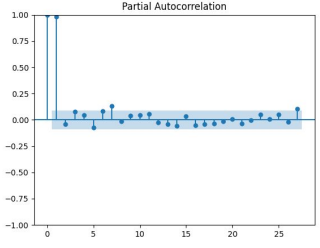
all



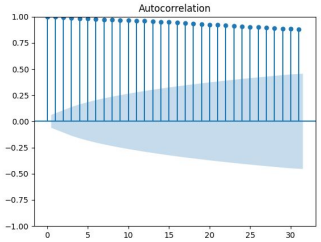
ACF and PACF plots:

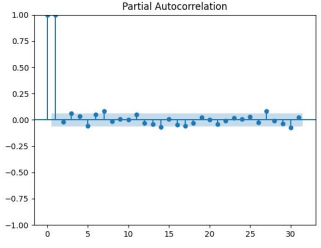
Autocorrelation function = correlation of the time series with its lags; Partial autocorrelation function = partial correlation of time series with its lags after removing the effects of lower order lags between them (https://www.youtube.com/watch?v=DeORzP0go5I&t=1s)



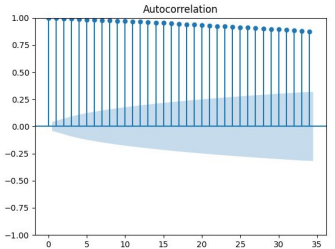


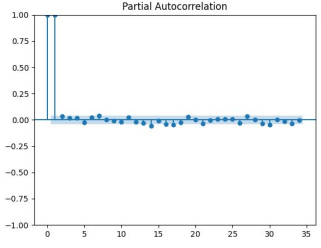
3 years



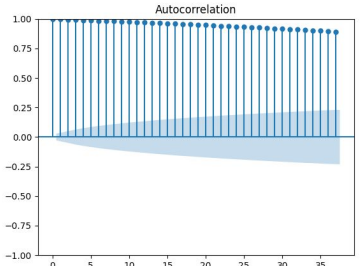


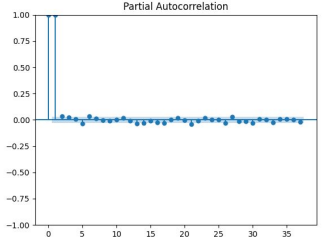
5 years



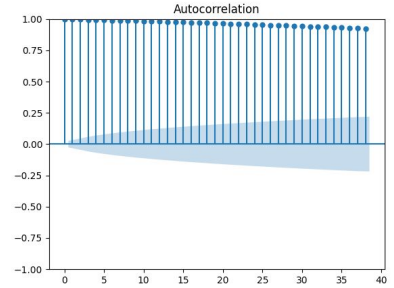


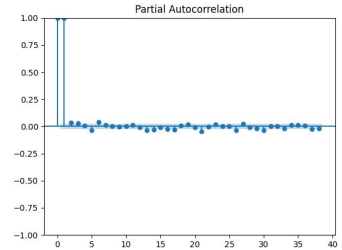
10 years





20 years





all

The ACF plot show that the correlations with the lags are high and positive with a slow decline. The PACF plot show a spike at 1 and then small or no spikes afterwards. These are signs of a simple random walk, a common time series, which is not stationary.

ADF Test:

Augmented Dickey-Fuller Test = tests for stationarity. The null hypothesis is that the data has a unit root and is non-stationary.

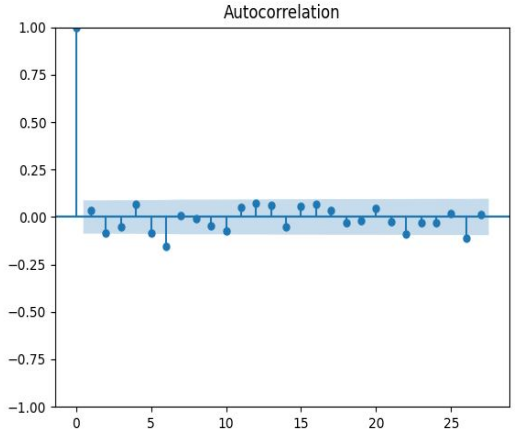
The results for each of the subsets of data:

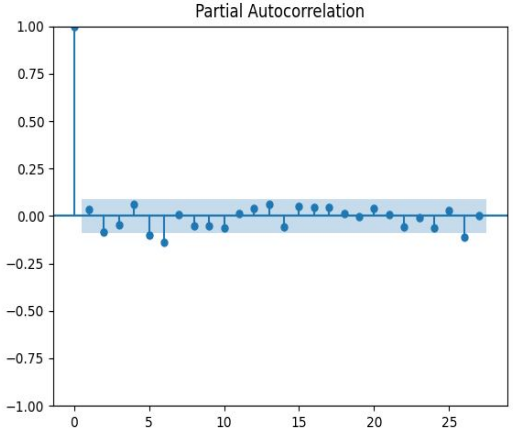
- Past 3 years = 0.53 ⇒ fail to reject null hypothesis (data is not stationary)
- Past 5 years = 0.63 ⇒ fail to reject null hypothesis (data is not stationary)
- Past 10 years = 0.17 ⇒ fail to reject null hypothesis (data in not stationary)
- Past 20 years = 0.03 ⇒ can reject null hypothesis (data seems to be stationary)
- All years = 0.08 ⇒ fail to reject null hypothesis (data is not stationary)

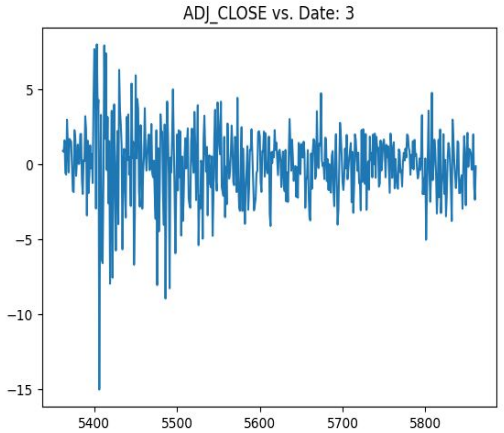
I will difference all data, and for the 20 years I will see both the behavior of differencing and not differencing (although it seems like it should still be differenced because the ACF and PACF do not indicate that it is stationary)

Differenced Data: 3 years

After one differencing it is clearer that the differenced data is more stationary. The p-value of the ADF test is 6.28e-23, so we can reject the null and confidently say that the differenced data is stationary.



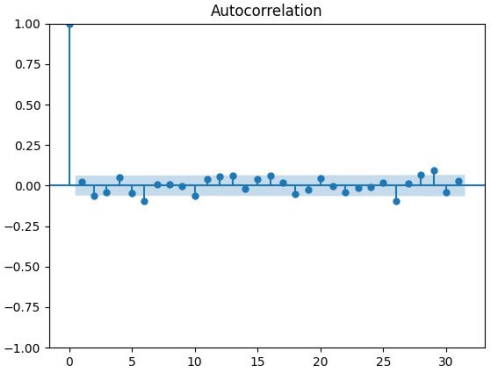


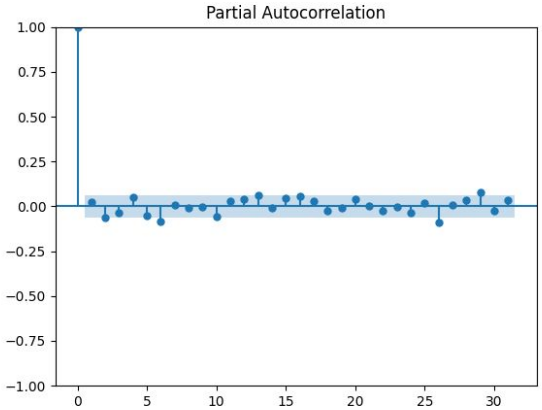


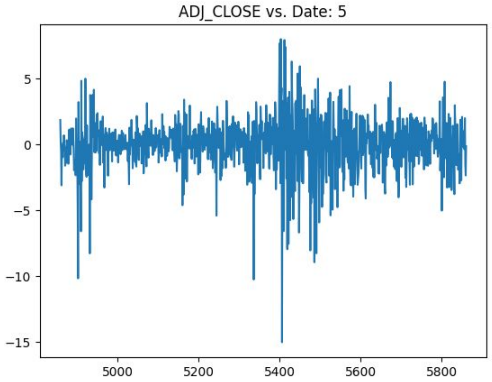
There is a significant spike at lag 6 in both the ACF and PACF plots. This implies p = 6/q = 0 and p = 0/q = 6, respectively. d = 1 because we only differenced once.

Differenced Data: 5 years

After one differencing it is clearer that the differenced data is more stationary. The p-value of the ADF test is 1.26e-20, so we can reject the null and confidently say that the differenced data is stationary.



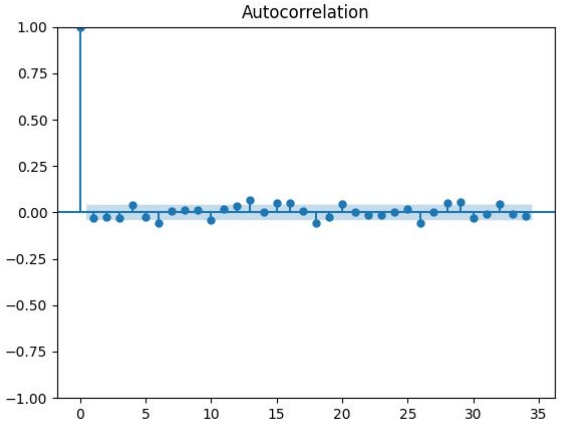


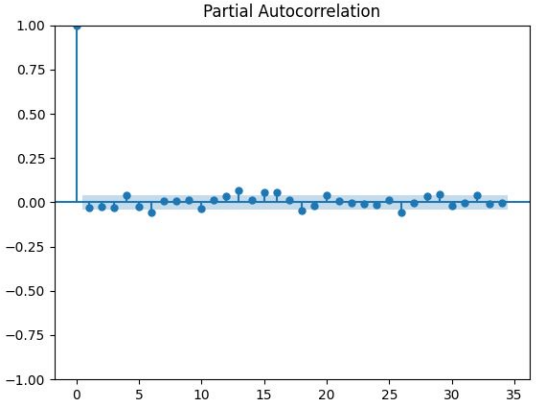


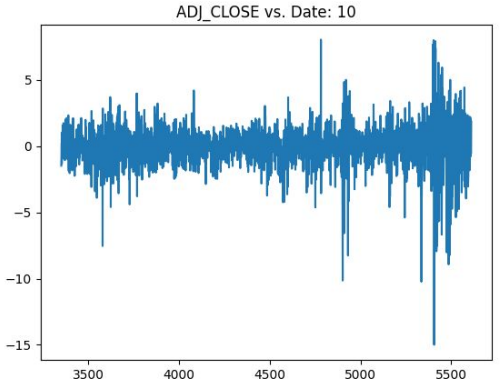
There is a significant spike at lag 6 in both the ACF and PACF plots. This implies p = 6/q = 0 and p = 0/q = 6, respectively. d = 1 because we only differenced once.

Differenced Data: 10 years

After one differencing it is clearer that the differenced data is more stationary. The p-value of the ADF test is 1.71e-15, so we can reject the null and confidently say that the differenced data is stationary.



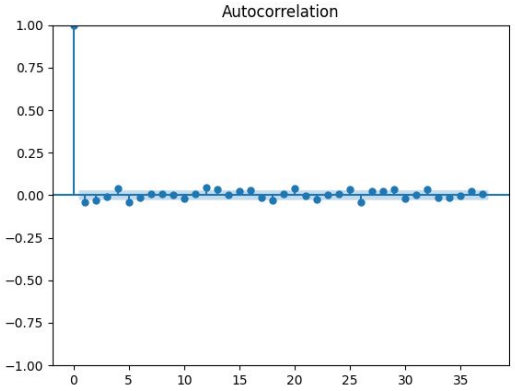


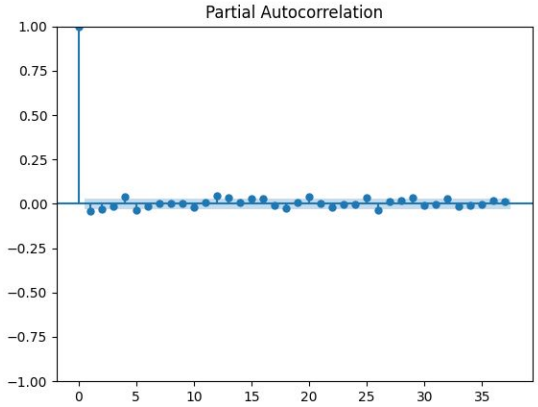


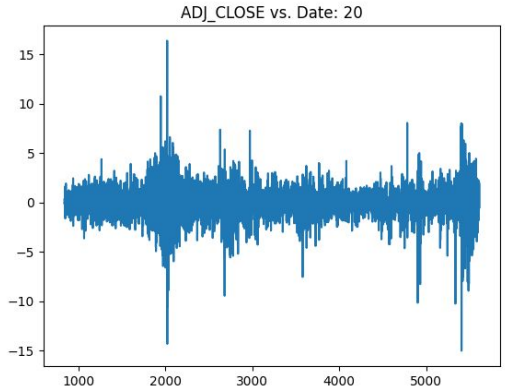
There is a significant spike at lag 6 in both the ACF and PACF plots. This implies p = 6/q = 0 and p = 0/q = 6, respectively. d = 1 because we only differenced once.

Differenced Data: 20 years

After one differencing it is clearer that the differenced data is more stationary. The p-value of the ADF test is 2.94e-27, so we can reject the null and confidently say that the differenced data is stationary.



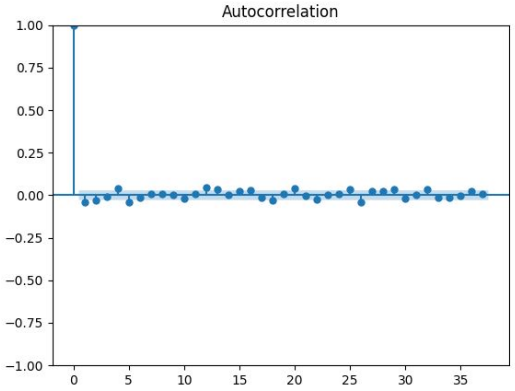


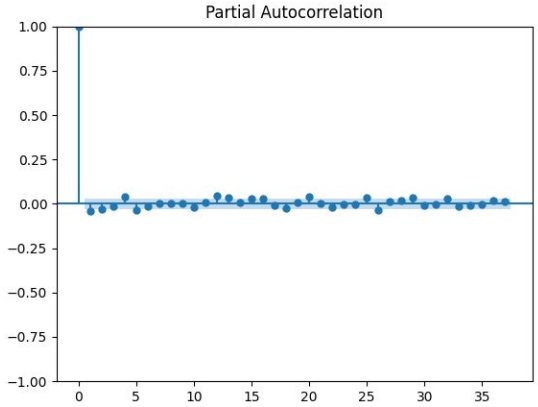


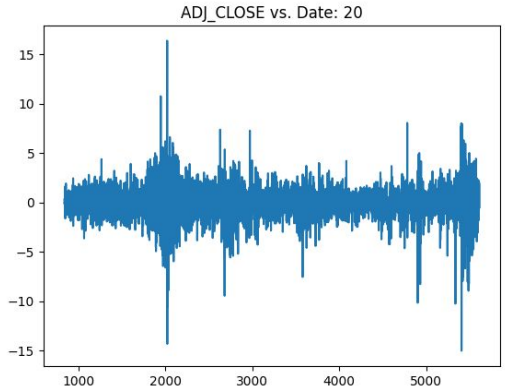
There is a significant spike at lag 4 in both the ACF and PACF plots. This implies p = 4/q = 0 and p = 0/q = 4, respectively. d = 1 because we only differenced once.

Differenced Data: Full dataset

After one differencing it is clearer that the differenced data is more stationary. The p-value of the ADF test is 1.19e-20, so we can reject the null and confidently say that the differenced data is stationary.







There is a significant spike at lag 4 in both the ACF and PACF plots. This implies p = 4/q = 0 and p = 0/q = 4, respectively. d = 1 because we only differenced once.

Results

Methods

3 types of models

- Auto generated parameters
- Only AR and I parameters
- Only MA and I parameters

Comparing the performance of each type of model across the different sized subsets of data.

Auto generated parameters: full data

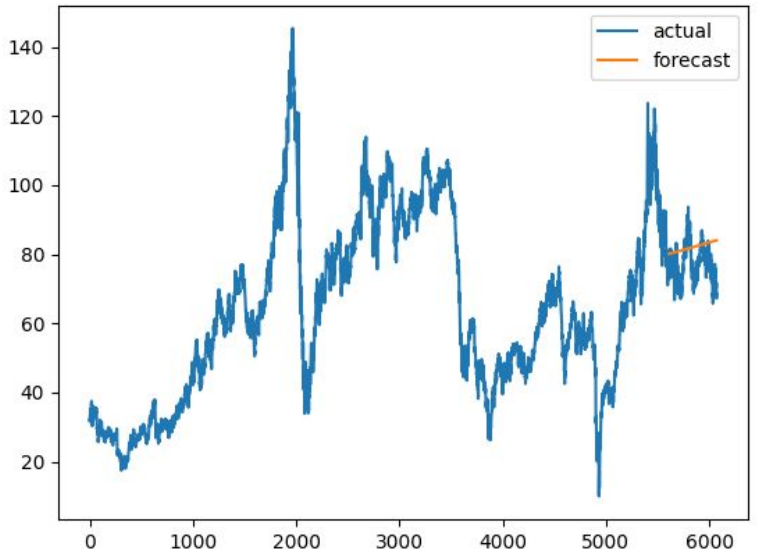
SARIMAX Results

Dep. Variable:	y	No. Observations:	5612
Model:	SARIMAX(0, 1, 5)	Log Likelihood	-10387.213
Date:	Wed, 07 May 2025	AIC	20788.426
Time:	11:57:29	BIC	20834.854
Sample:	0 - 5612	HQIC	20804.604
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	0.0086	0.019	0.443	0.658	-0.029	0.047
ma.L1	-0.0377	0.007	-5.612	0.000	-0.051	-0.025
ma.L2	-0.0266	0.008	-3.171	0.002	-0.043	-0.010
ma.L3	-0.0081	0.008	-1.025	0.305	-0.024	0.007
ma.L4	0.0366	0.008	4.811	0.000	0.022	0.051
ma.L5	-0.0423	0.007	-5.717	0.000	-0.057	-0.028
sigma2	2.3739	0.020	119.259	0.000	2.335	2.413

Ljung-Box (L1) (Q):	0.00	Jarque-Bera (JB):	20437.73
Prob(Q):	0.97	Prob(JB):	0.00
Heteroskedasticity (H):	2.55	Skew:	-0.37
Prob(H) (two-sided):	0.00	Kurtosis:	12.32

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).



mean absolute error: 6.0411036512809755
mean absolute percentage error: 0.0814159644326747
residual mean squared error: 7.3642166352007425

Auto generated parameters: 20 years

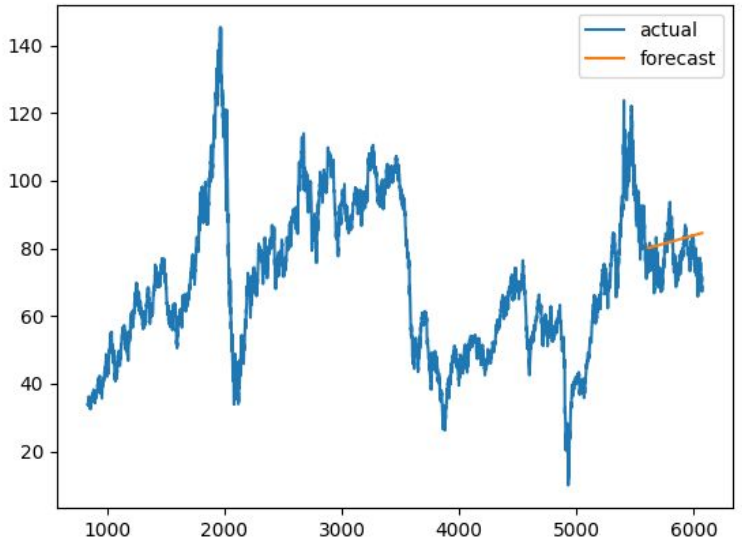
SARIMAX Results

Dep. Variable:	y	No. Observations:	4776			
Model:	SARIMAX(0, 1, 5)	Log Likelihood	-9147.135			
Date:	Wed, 07 May 2025	AIC	18308.270			
Time:	11:57:53	BIC	18353.568			
Sample:	0 - 4776	HQIC	18324.184			
Covariance Type:	opg					

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	0.0097	0.022	0.433	0.665	-0.034	0.054
ma.L1	-0.0392	0.008	-5.119	0.000	-0.054	-0.024
ma.L2	-0.0265	0.010	-2.767	0.006	-0.045	-0.008
ma.L3	-0.0097	0.009	-1.086	0.278	-0.027	0.008
ma.L4	0.0384	0.009	4.423	0.000	0.021	0.055
ma.L5	-0.0420	0.008	-4.976	0.000	-0.058	-0.025
sigma2	2.7003	0.026	104.080	0.000	2.649	2.751

Ljung-Box (L1) (Q):	0.00	Jarque-Bera (JB):	13322.27
Prob(Q):	0.97	Prob(JB):	0.00
Heteroskedasticity (H):	0.90	Skew:	-0.34
Prob(H) (two-sided):	0.04	Kurtosis:	11.15

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).



mean absolute error: 6.216849279120286
mean absolute percentage error: 0.08380943795302777
residual mean squared error: 7.558289320068495

Auto generated parameters: 10 years

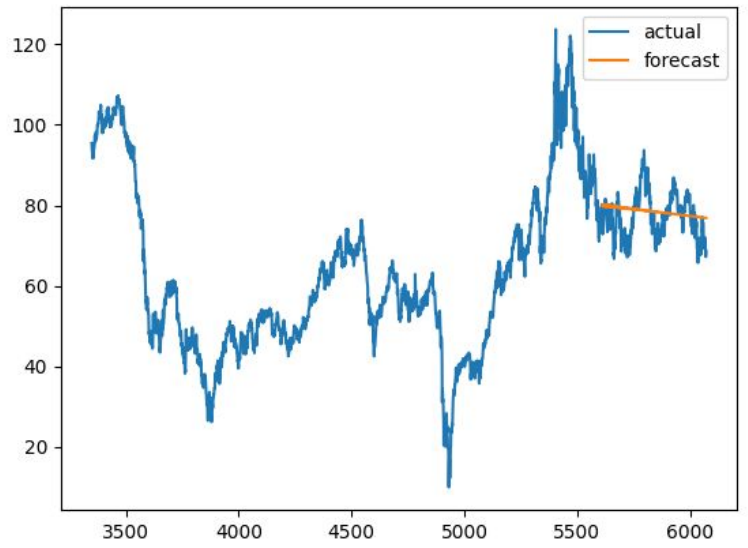
SARIMAX Results

Dep. Variable:	y	No. Observations:	2263			
Model:	SARIMAX(2, 1, 2)	Log Likelihood	-4229.805			
Date:	Wed, 07 May 2025	AIC	8471.610			
Time:	11:58:11	BIC	8505.954			
Sample:	0	HQIC	8484.142			
	- 2263					
Covariance Type:	opg					

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	-0.0133	0.066	-0.201	0.840	-0.143	0.116
ar.L1	0.0416	0.007	5.877	0.000	0.028	0.055
ar.L2	-0.9839	0.007	-138.059	0.000	-0.998	-0.970
ma.L1	-0.0476	0.010	-4.664	0.000	-0.068	-0.028
ma.L2	0.9628	0.010	96.732	0.000	0.943	0.982
sigma2	2.4636	0.033	74.508	0.000	2.399	2.528

Ljung-Box (L1) (Q):	1.41	Jarque-Bera (JB):	7488.81
Prob(Q):	0.23	Prob(JB):	0.00
Heteroskedasticity (H):	2.77	Skew:	-0.84
Prob(H) (two-sided):	0.00	Kurtosis:	11.76

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).



mean absolute error: 4.570801424542164
mean absolute percentage error: 0.06005604886165142
residual mean squared error: 5.633087283700272

Auto generated parameters: 5 years

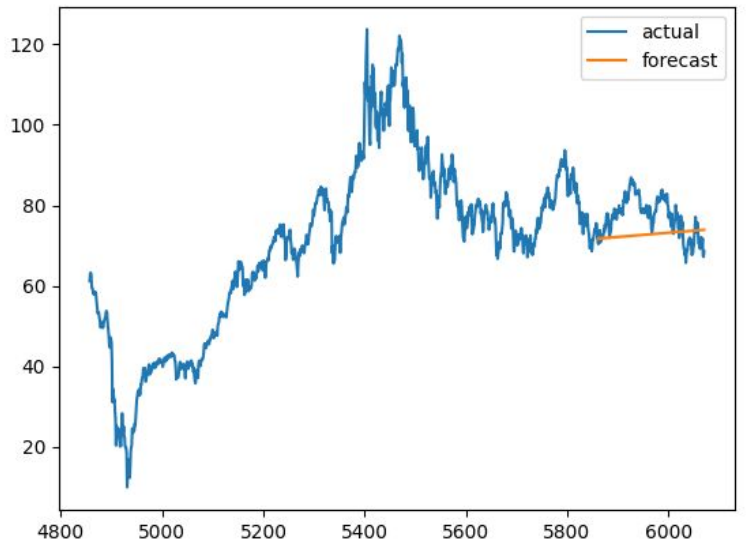
SARIMAX Results

Dep. Variable:	y	No. Observations:	1005			
Model:	SARIMAX(2, 1, 2)	Log Likelihood	-2133.346			
Date:	Wed, 07 May 2025	AIC	4278.691			
Time:	11:58:21	BIC	4308.162			
Sample:	0 - 1005	HQIC	4289.890			
Covariance Type:	opg					

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	0.0202	0.128	0.157	0.875	-0.231	0.272
ar.L1	0.0441	0.013	3.273	0.001	0.018	0.071
ar.L2	-0.9758	0.014	-68.180	0.000	-1.004	-0.948
ma.L1	-0.0513	0.020	-2.571	0.010	-0.090	-0.012
ma.L2	0.9436	0.020	47.023	0.000	0.904	0.983
sigma2	4.1018	0.099	41.541	0.000	3.908	4.295

Ljung-Box (L1) (Q):	0.86	Jarque-Bera (JB):	1488.55
Prob(Q):	0.36	Prob(JB):	0.00
Heteroskedasticity (H):	1.25	Skew:	-0.84
Prob(H) (two-sided):	0.04	Kurtosis:	8.72

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).



mean absolute error: 5.530824841047598
mean absolute percentage error: 0.06996323162062335
residual mean squared error: 6.546903792567009

Auto generated parameters: 3 years

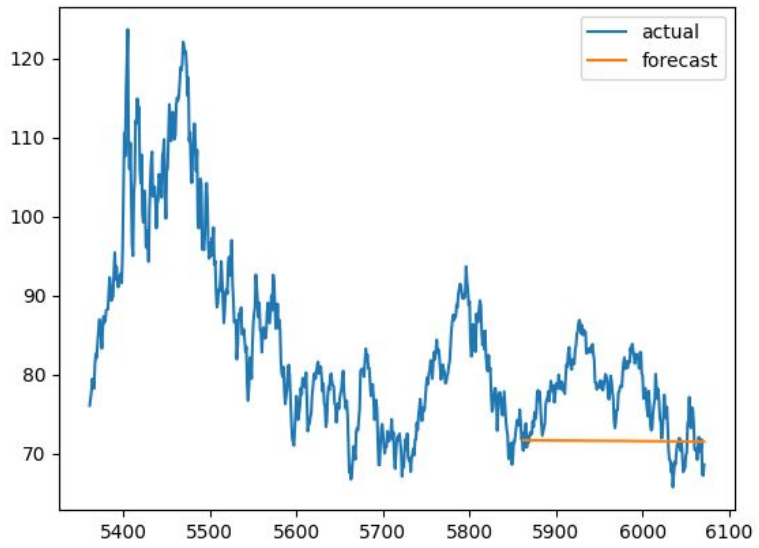
SARIMAX Results

Dep. Variable:	y	No. Observations:	501
Model:	SARIMAX(2, 1, 2)	Log Likelihood	-1146.495
Date:	Wed, 07 May 2025	AIC	2304.991
Time:	11:58:29	BIC	2330.279
Sample:	0 - 501	HQIC	2314.914
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	-0.0016	0.215	-0.007	0.994	-0.422	0.419
ar.L1	0.0351	0.020	1.754	0.079	-0.004	0.074
ar.L2	-0.9772	0.019	-51.428	0.000	-1.014	-0.940
ma.L1	-0.0502	0.029	-1.717	0.086	-0.108	0.007
ma.L2	0.9562	0.028	33.959	0.000	0.901	1.011
sigma2	5.7712	0.224	25.752	0.000	5.332	6.210

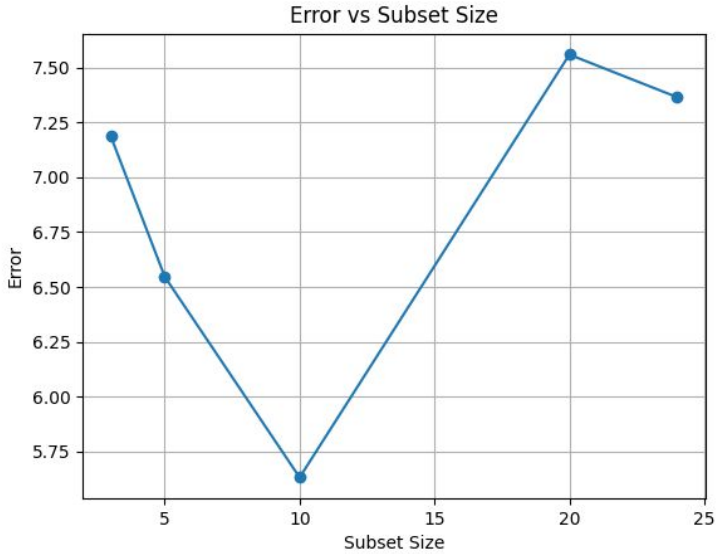
Ljung-Box (L1) (Q):	1.44	Jarque-Bera (JB):	323.76
Prob(Q):	0.23	Prob(JB):	0.00
Heteroskedasticity (H):	0.24	Skew:	-0.56
Prob(H) (two-sided):	0.00	Kurtosis:	6.78

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).



mean absolute error: 6.045010860403555
mean absolute percentage error: 0.07595413066603919
residual mean squared error: 7.186302468294149

Auto generated parameters: error vs subset size



10 years of data has the best residual mean squared error.

Only AR parameter: full data

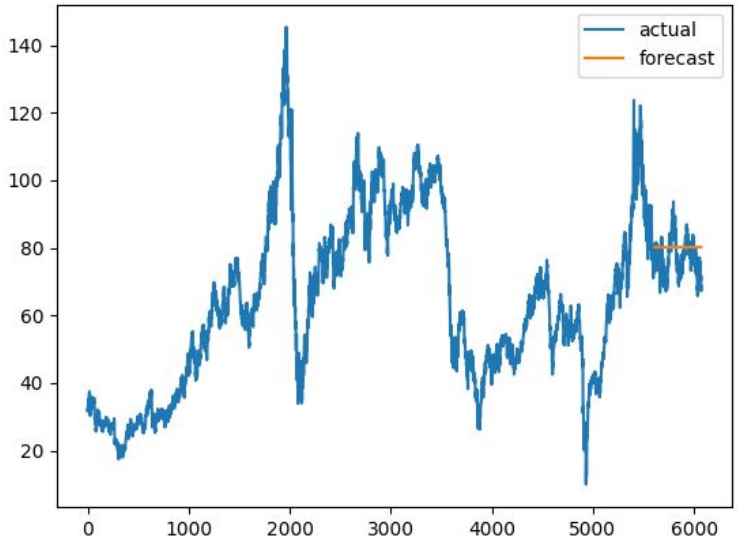
SARIMAX Results

Dep. Variable:	adj_close	No. Observations:	5612
Model:	ARIMA(4, 1, 0)	Log Likelihood	-10391.573
Date:	Wed, 07 May 2025	AIC	20793.145
Time:	11:57:30	BIC	20826.308
Sample:	0 - 5612	HQIC	20804.701
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.0380	0.007	-5.735	0.000	-0.051	-0.025
ar.L2	-0.0291	0.008	-3.545	0.000	-0.045	-0.013
ar.L3	-0.0087	0.008	-1.138	0.255	-0.024	0.006
ar.L4	0.0376	0.007	5.174	0.000	0.023	0.052
sigma2	2.3777	0.019	126.515	0.000	2.341	2.414

Ljung-Box (L1) (Q):	0.01	Jarque-Bera (JB):	21488.91
Prob(Q):	0.92	Prob(JB):	0.00
Heteroskedasticity (H):	2.56	Skew:	-0.36
Prob(H) (two-sided):	0.00	Kurtosis:	12.56

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).



mean absolute error: 5.048268230105487
mean absolute percentage error: 0.06746413019293145
residual mean squared error: 6.182928904804433

Only AR parameters: 20 years

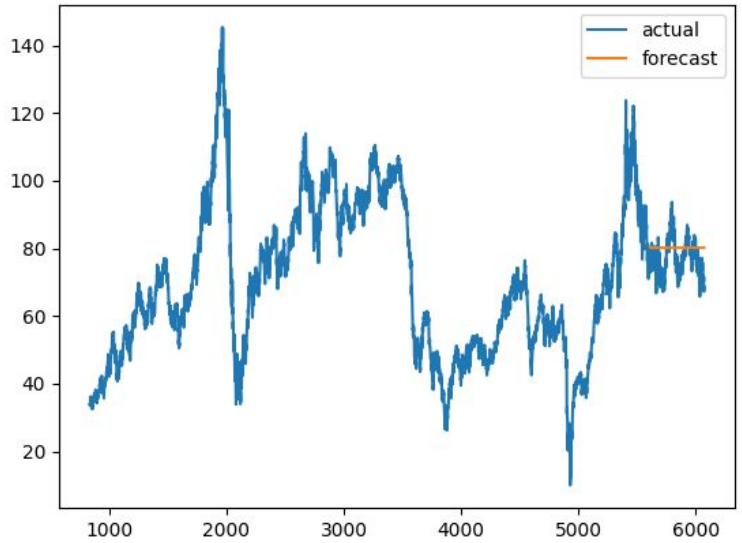
SARIMAX Results

Dep. Variable:	adj_close	No. Observations:	4776
Model:	ARIMA(4, 1, 0)	Log Likelihood	-9150.755
Date:	Wed, 07 May 2025	AIC	18311.509
Time:	11:57:54	BIC	18343.865
Sample:	0 - 4776	HQIC	18322.876
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.0396	0.008	-5.256	0.000	-0.054	-0.025
ar.L2	-0.0290	0.009	-3.102	0.002	-0.047	-0.011
ar.L3	-0.0105	0.009	-1.205	0.228	-0.028	0.007
ar.L4	0.0392	0.008	4.730	0.000	0.023	0.055
sigma2	2.7044	0.024	110.519	0.000	2.656	2.752

Ljung-Box (L1) (Q):	0.01	Jarque-Bera (JB):	14025.73
Prob(Q):	0.92	Prob(JB):	0.00
Heteroskedasticity (H):	0.90	Skew:	-0.34
Prob(H) (two-sided):	0.03	Kurtosis:	11.37

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).



mean absolute error: 5.047081085128184
mean absolute percentage error: 0.0674687959669475
residual mean squared error: 6.181587396291245

Only AR parameters: 10 years

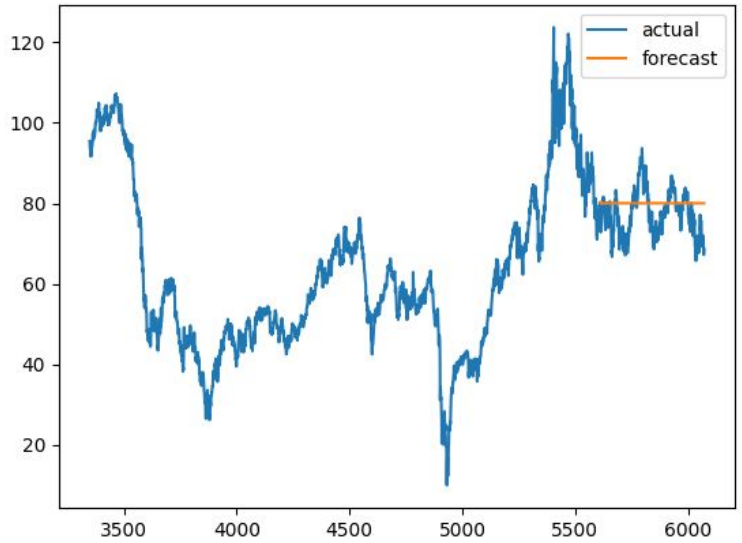
SARIMAX Results

Dep. Variable:	adj_close	No. Observations:	2263
Model:	ARIMA(6, 1, 0)	Log Likelihood	-4238.190
Date:	Wed, 07 May 2025	AIC	8490.380
Time:	11:58:12	BIC	8530.448
Sample:	0 - 2263	HQIC	8505.000
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.0302	0.014	-2.237	0.025	-0.057	-0.004
ar.L2	-0.0234	0.013	-1.814	0.070	-0.049	0.002
ar.L3	-0.0305	0.011	-2.654	0.008	-0.053	-0.008
ar.L4	0.0373	0.013	2.943	0.003	0.012	0.062
ar.L5	-0.0238	0.013	-1.892	0.059	-0.049	0.001
ar.L6	-0.0567	0.013	-4.407	0.000	-0.082	-0.031
sigma2	2.4828	0.036	69.081	0.000	2.412	2.553

Ljung-Box (L1) (Q):	0.00	Jarque-Bera (JB):	6420.36
Prob(Q):	0.99	Prob(JB):	0.00
Heteroskedasticity (H):	2.82	Skew:	-0.82
Prob(H) (two-sided):	0.00	Kurtosis:	11.09

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).



mean absolute error: 4.992347621530395
mean absolute percentage error: 0.06664481722083114
residual mean squared error: 6.117747640251688

Only AR parameters: 5 years

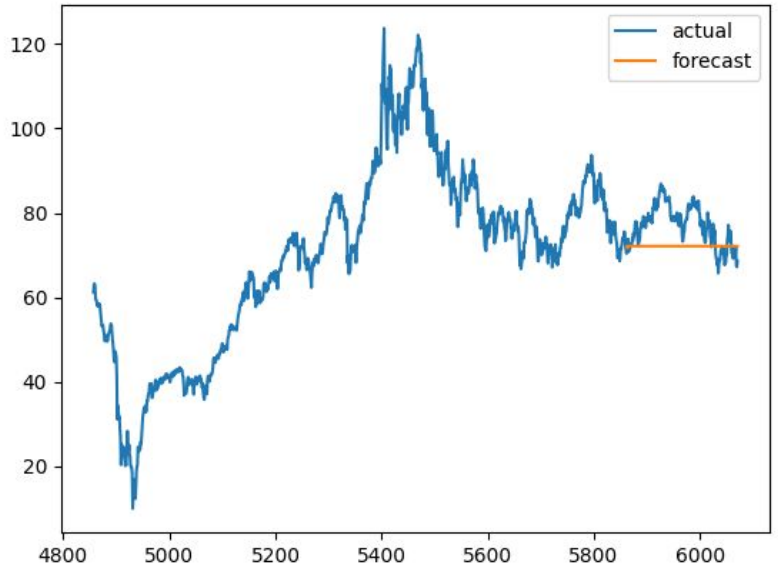
SARIMAX Results

Dep. Variable:	adj_close	No. Observations:	1005
Model:	ARIMA(6, 1, 0)	Log Likelihood	-2135.775
Date:	Wed, 07 May 2025	AIC	4285.549
Time:	11:58:22	BIC	4319.931
Sample:	0 - 1005	HQIC	4298.614
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.0200	0.024	0.821	0.412	-0.028	0.068
ar.L2	-0.0584	0.022	-2.638	0.008	-0.102	-0.015
ar.L3	-0.0458	0.019	-2.358	0.018	-0.084	-0.008
ar.L4	0.0458	0.022	2.075	0.038	0.003	0.089
ar.L5	-0.0522	0.022	-2.429	0.015	-0.094	-0.010
ar.L6	-0.0864	0.023	-3.833	0.000	-0.131	-0.042
sigma2	4.1229	0.107	38.651	0.000	3.914	4.332

Ljung-Box (L1) (Q):	0.00	Jarque-Bera (JB):	1212.42
Prob(Q):	0.98	Prob(JB):	0.00
Heteroskedasticity (H):	1.20	Skew:	-0.83
Prob(H) (two-sided):	0.10	Kurtosis:	8.12

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).



mean absolute error: 5.710568337699619
mean absolute percentage error: 0.07183979191057999
residual mean squared error: 6.810863986801155

Only AR parameters: 3 years

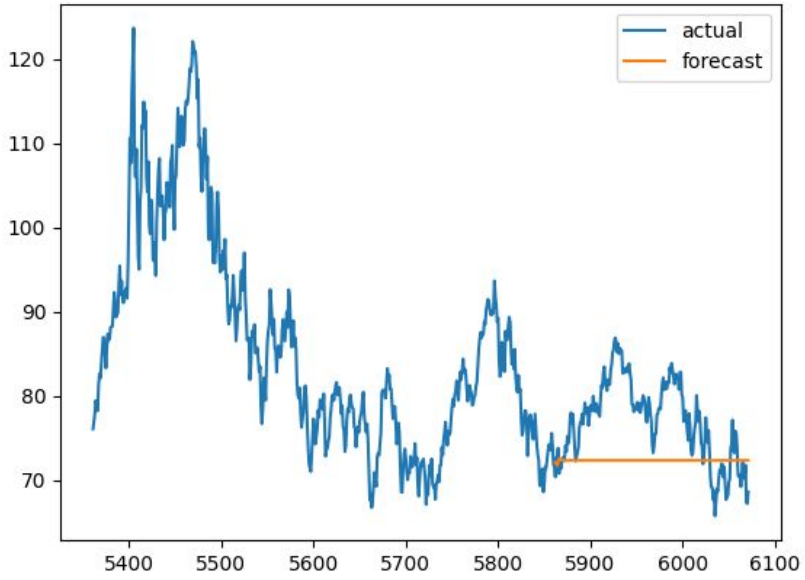
SARIMAX Results

Dep. Variable:	adj_close	No. Observations:	501
Model:	ARIMA(6, 1, 0)	Log Likelihood	-1143.880
Date:	Wed, 07 May 2025	AIC	2301.761
Time:	11:58:29	BIC	2331.263
Sample:	0 - 501	HQIC	2313.338
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.0309	0.041	0.748	0.454	-0.050	0.112
ar.L2	-0.0736	0.035	-2.129	0.033	-0.141	-0.006
ar.L3	-0.0643	0.028	-2.274	0.023	-0.120	-0.009
ar.L4	0.0562	0.033	1.707	0.088	-0.008	0.121
ar.L5	-0.0927	0.032	-2.887	0.004	-0.156	-0.030
ar.L6	-0.1402	0.037	-3.784	0.000	-0.213	-0.068
sigma2	5.6815	0.267	21.290	0.000	5.158	6.205

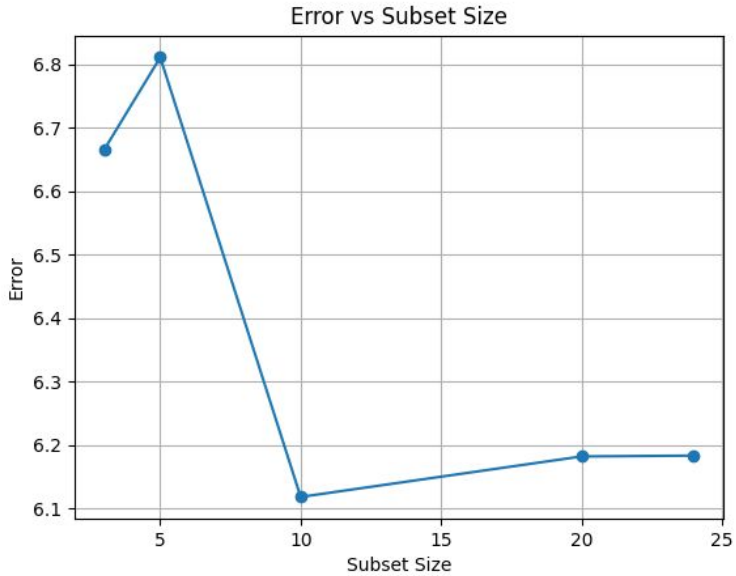
Ljung-Box (L1) (Q):	0.00	Jarque-Bera (JB):	190.45
Prob(Q):	0.97	Prob(JB):	0.00
Heteroskedasticity (H):	0.24	Skew:	-0.45
Prob(H) (two-sided):	0.00	Kurtosis:	5.88

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).



mean absolute error: 5.585222184360733
mean absolute percentage error: 0.07030817536113423
residual mean squared error: 6.665028566557918

Only AR parameters: error vs subset size



10 years of data has the best residual mean squared error, but not by a lot compared to the other subset sizes

Only MA parameter: full data

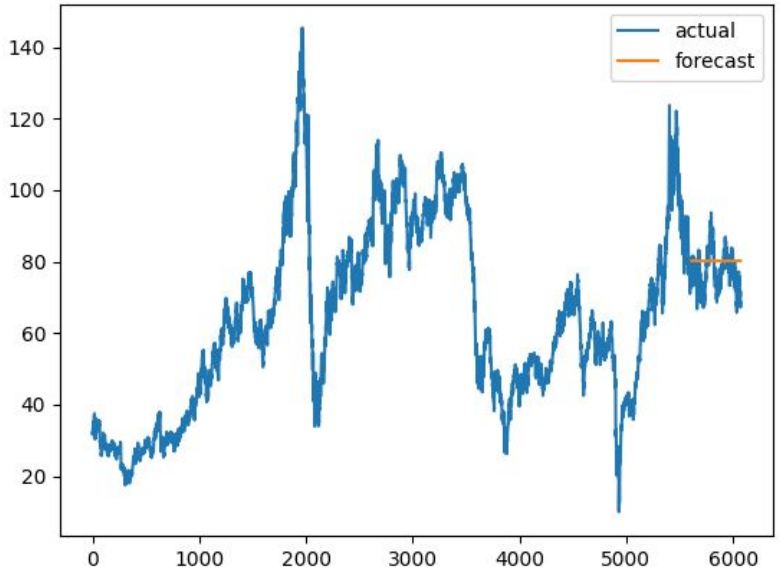
SARIMAX Results

Dep. Variable:	adj_close	No. Observations:	5612
Model:	ARIMA(0, 1, 4)	Log Likelihood	-10392.232
Date:	Wed, 07 May 2025	AIC	20794.464
Time:	11:57:31	BIC	20827.626
Sample:	0 - 5612	HQIC	20806.019
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ma.L1	-0.0351	0.007	-5.349	0.000	-0.048	-0.022
ma.L2	-0.0276	0.008	-3.292	0.001	-0.044	-0.011
ma.L3	-0.0095	0.008	-1.221	0.222	-0.025	0.006
ma.L4	0.0354	0.007	4.771	0.000	0.021	0.050
sigma2	2.3782	0.019	126.466	0.000	2.341	2.415

Ljung-Box (L1) (Q):	0.01	Jarque-Bera (JB):	21555.56
Prob(Q):	0.91	Prob(JB):	0.00
Heteroskedasticity (H):	2.56	Skew:	-0.36
Prob(H) (two-sided):	0.00	Kurtosis:	12.57

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).



mean absolute error: 5.051131467142908
mean absolute percentage error: 0.0675056978697396 3
residual mean squared error: 6.186154286025030 5

Only MA parameter: 20 years

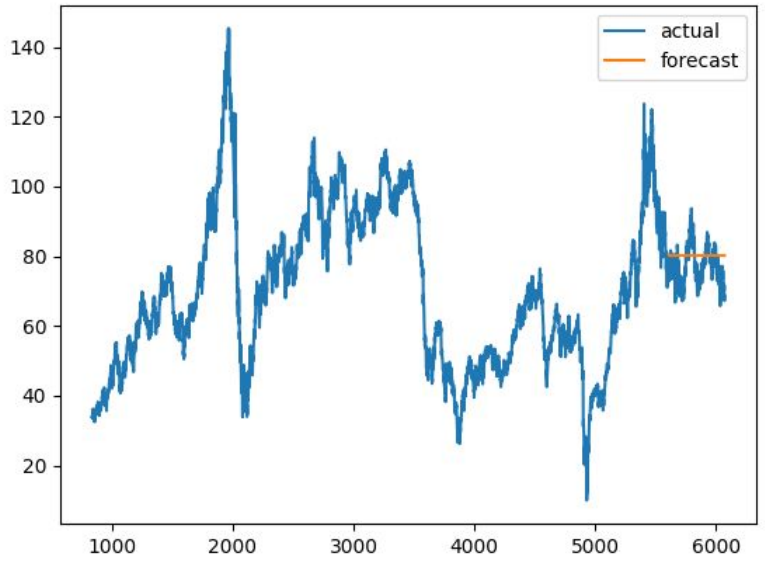
SARIMAX Results

Dep. Variable:	adj_close	No. Observations:	4776
Model:	ARIMA(0, 1, 4)	Log Likelihood	-9151.334
Date:	Wed, 07 May 2025	AIC	18312.668
Time:	11:57:55	BIC	18345.024
Sample:	0 - 4776	HQIC	18324.035
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ma.L1	-0.0366	0.007	-4.909	0.000	-0.051	-0.022
ma.L2	-0.0275	0.010	-2.882	0.004	-0.046	-0.009
ma.L3	-0.0112	0.009	-1.270	0.204	-0.028	0.006
ma.L4	0.0370	0.008	4.377	0.000	0.020	0.054
sigma2	2.7051	0.024	110.457	0.000	2.657	2.753

Ljung-Box (L1) (Q):	0.01	Jarque-Bera (JB):	14072.90
Prob(Q):	0.91	Prob(JB):	0.00
Heteroskedasticity (H):	0.90	Skew:	-0.34
Prob(H) (two-sided):	0.03	Kurtosis:	11.38

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).



mean absolute error: 5.050123031896907 5
mean absolute percentage error: 0.06749105979849974
residual mean squared error: 6.185020670714847

Only MA parameter: 10 years

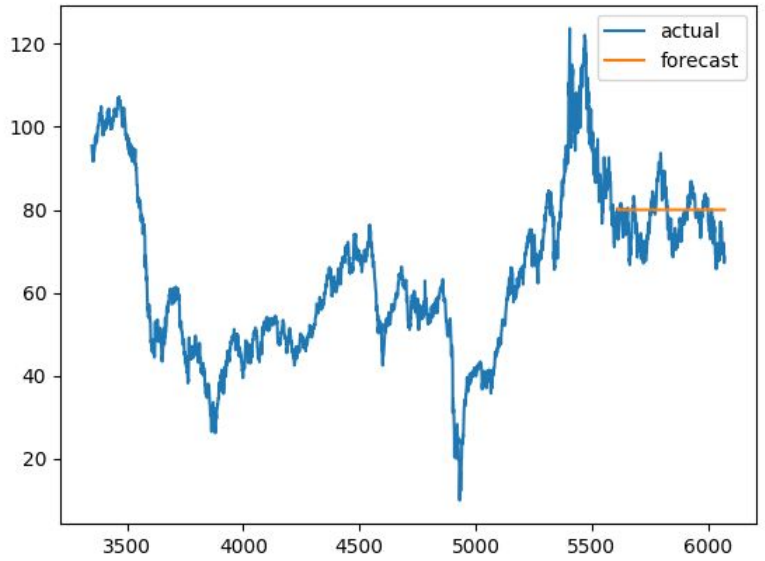
SARIMAX Results

Dep. Variable:	adj_close	No. Observations:	2263
Model:	ARIMA(0, 1, 6)	Log Likelihood	-4238.989
Date:	Wed, 07 May 2025	AIC	8491.977
Time:	11:58:13	BIC	8532.045
Sample:	0	HQIC	8506.598
	- 2263		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ma.L1	-0.0289	0.014	-2.140	0.032	-0.055	-0.002
ma.L2	-0.0208	0.013	-1.596	0.110	-0.046	0.005
ma.L3	-0.0278	0.012	-2.323	0.020	-0.051	-0.004
ma.L4	0.0375	0.013	2.847	0.004	0.012	0.063
ma.L5	-0.0242	0.012	-1.945	0.052	-0.049	0.000
ma.L6	-0.0475	0.013	-3.599	0.000	-0.073	-0.022
sigma2	2.4846	0.036	69.139	0.000	2.414	2.555

Ljung-Box (L1) (Q):	0.00	Jarque-Bera (JB):	6519.61
Prob(Q):	0.97	Prob(JB):	0.00
Heteroskedasticity (H):	2.82	Skew:	-0.82
Prob(H) (two-sided):	0.00	Kurtosis:	11.15

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).



mean absolute error: 4.983434026759943 5
mean absolute percentage error: 0.06651331369804896
residual mean squared error: 6.107166987098335

Only MA parameter: 5 years

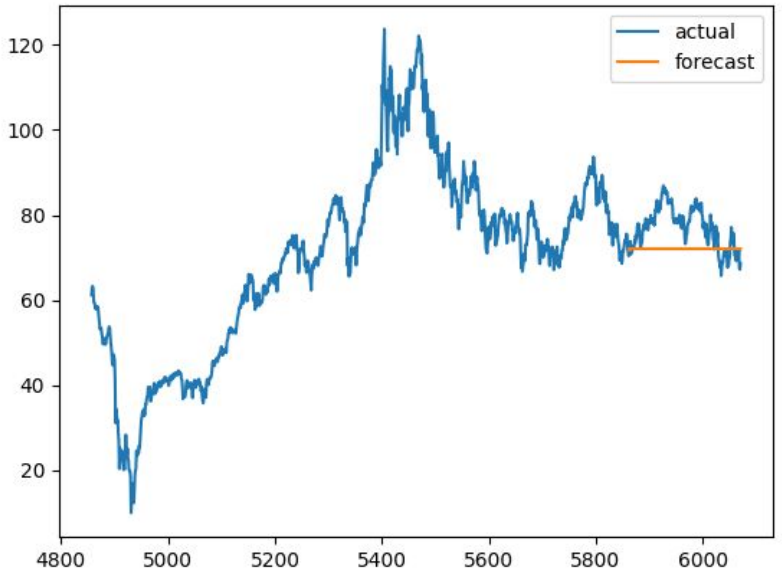
SARIMAX Results

Dep. Variable:	adj_close	No. Observations:	1005
Model:	ARIMA(0, 1, 6)	Log Likelihood	-2136.603
Date:	Wed, 07 May 2025	AIC	4287.207
Time:	11:58:23	BIC	4321.589
Sample:	0	HQIC	4300.272
	- 1005		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ma.L1	0.0221	0.024	0.915	0.360	-0.025	0.070
ma.L2	-0.0576	0.022	-2.591	0.010	-0.101	-0.014
ma.L3	-0.0482	0.020	-2.419	0.016	-0.087	-0.009
ma.L4	0.0427	0.023	1.826	0.068	-0.003	0.089
ma.L5	-0.0409	0.021	-1.930	0.054	-0.083	0.001
ma.L6	-0.0781	0.023	-3.393	0.001	-0.123	-0.033
sigma2	4.1298	0.107	38.591	0.000	3.920	4.340

Ljung-Box (L1) (Q):	0.00	Jarque-Bera (JB):	1246.01
Prob(Q):	0.97	Prob(JB):	0.00
Heteroskedasticity (H):	1.20	Skew:	-0.85
Prob(H) (two-sided):	0.09	Kurtosis:	8.19

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).



mean absolute error: 5.743960011185369
mean absolute percentage error: 0.07224958389212578
residual mean squared error: 6.84883727234121

Only MA parameter: 3 years

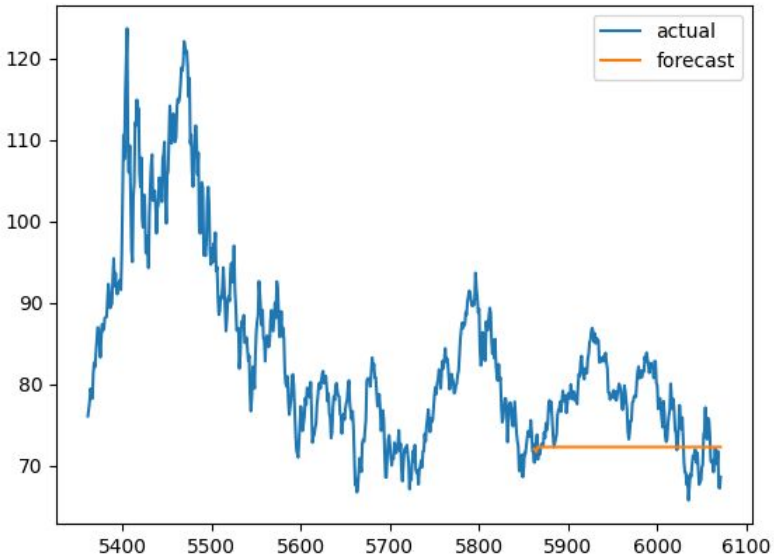
SARIMAX Results

Dep. Variable:	adj_close	No. Observations:	501
Model:	ARIMA(0, 1, 6)	Log Likelihood	-1144.102
Date:	Wed, 07 May 2025	AIC	2302.204
Time:	11:58:31	BIC	2331.706
Sample:	0 - 501	HQIC	2313.781
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ma.L1	0.0346	0.041	0.845	0.398	-0.046	0.115
ma.L2	-0.0818	0.035	-2.359	0.018	-0.150	-0.014
ma.L3	-0.0869	0.030	-2.940	0.003	-0.145	-0.029
ma.L4	0.0478	0.037	1.286	0.198	-0.025	0.121
ma.L5	-0.0712	0.032	-2.246	0.025	-0.133	-0.009
ma.L6	-0.1373	0.037	-3.677	0.000	-0.210	-0.064
sigma2	5.6868	0.264	21.503	0.000	5.168	6.205

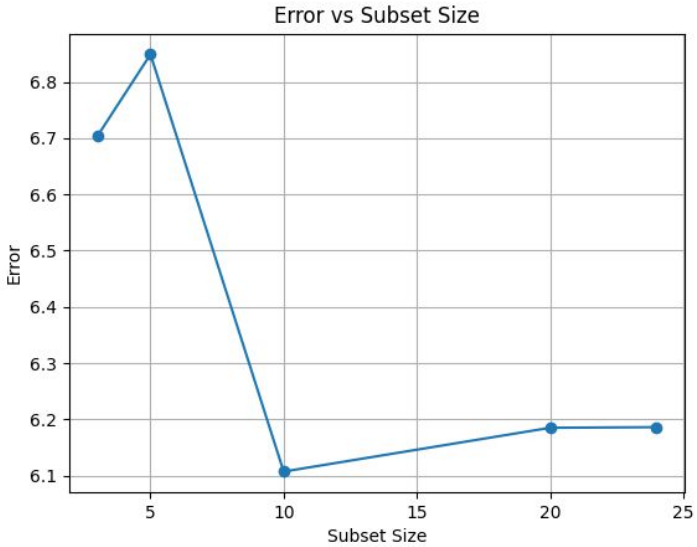
Ljung-Box (L1) (Q):	0.00	Jarque-Bera (JB):	194.29
Prob(Q):	0.96	Prob(JB):	0.00
Heteroskedasticity (H):	0.24	Skew:	-0.46
Prob(H) (two-sided):	0.00	Kurtosis:	5.91

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).



mean absolute error: 5.616602821308457
mean absolute percentage error: 0.0706858609460543
residual mean squared error: 6.702965772731589

Only MA parameters: error vs subset size



10 years of data has the best residual mean squared error, but not by a lot compared to the other subset sizes

Takeaways

- 10 years of data is the sweet spot for not having too much or too little data (based on the errors)
- Manually chosen parameters performed as well, if not better, than the auto-generated parameters.
- Predictions tend to be linear.