Forecasting the Price of Silver

Introduction

Project Aims

This project seeks to forecast the price of -silver based on the analysis of historical price data. As a proxy for the price of silver, the exchange-traded fund (ETF) SLV is utilized, with daily price data ranging from January 1, 2017 to December 31, 2021. In this study, models used for time series analysis are applied to the time series data.

Business Problem

For centuries, silver has been used as a primary component of currency and, today, remains popular not only as a precious metal but also as an important component in industry in items such as solar panels, conductors, and batteries. On the investment side, besides its role in emerging technologies, silver has garnered interest as a hedge against inflation of the dollar and as a safe haven during periods of economic recession. However, with the recent popularity of cryptocurrencies, the viability of silver as an investment has been called into question. This study seeks to use historical data to forecast the price of SLV over the next 12 months to determine whether silver remains a good investment.

Stakeholders

Parties interested in this report would include investors and speculators of physical silver and silver paper assets.

Data Source

Historical price data of the iShares Silver Trust (stock symbol: SLV) were obtained from [Yahoo! Finance] and range from 1/1/2017 to 12/31/2021.

Exploratory Data Analysis

Data Wrangling

The dataset contains 1517 rows detailing the date, open, close, high, low, and volume of SLV (Table 1). The summary statistics for these data are presented in Table 2. There are no missing rows in the dataset. In addition, the *Adj Close* column is

	Date	Open	High	Low	Close	Adj Close	Volume
0	2017-01-03	15.25	15.64	15.19	15.44	15.44	11033700
1	2017-01-04	15.58	15.66	15.49	15.58	15.58	6142900
2	2017-01-05	15.73	15.85	15.66	15.76	15.76	7785600
3	2017-01-06	15.57	15.70	15.55	15.64	15.64	6097900
4	2017-01-09	15.69	15.83	15.65	15.70	15.70	7288900

Table 1. The first five rows of the SLV dataset, obtained from Yahoo! Finance.

identical to the *Close* column; thus, the *Adj Close* column is eliminated. The *Date* column is set as the index, and the data are treated as daily data without additional adjustment.

	Open	High	Low	Close	Adj Close	Volume
count	1259.000000	1259.000000	1259.000000	1259.000000	1259.000000	1.259000e+03
mean	17.717601	17.841493	17.573614	17.711477	17.711477	1.802589e+07
std	3.822863	3.873093	3.746981	3.814378	3.814378	1.862329e+07
min	11.340000	11.440000	10.860000	11.210000	11.210000	2.423800e+06
25%	15.070000	15.165000	15.020000	15.080000	15.080000	7.065250e+06
50%	16.150000	16.260000	16.040001	16.150000	16.150000	1.211190e+07
75%	21.410000	21.585000	21.200001	21.405001	21.405001	2.300390e+07
max	27.760000	27.980000	26.570000	27.000000	27.000000	2.806150e+08

Table 2. Statistics for the SLV dataset used in this study.

Data Overview

A plot of SLV over the timeframe of the dataset is shown in Figure 1. Over the past five years, SLV has been rather range bound, trading in two ranges—from 12 to 18 during 2017 to mid-2020, and from 20 to 27 from the third quarter of 2020 to the end of 2021. Though its performance during most of the study period can be characterized as fluctuation within ranges, there have been periods of high volatility, most notably the move from a low of 11.21 in March 2020 to a high of 27.00 in August 2020.

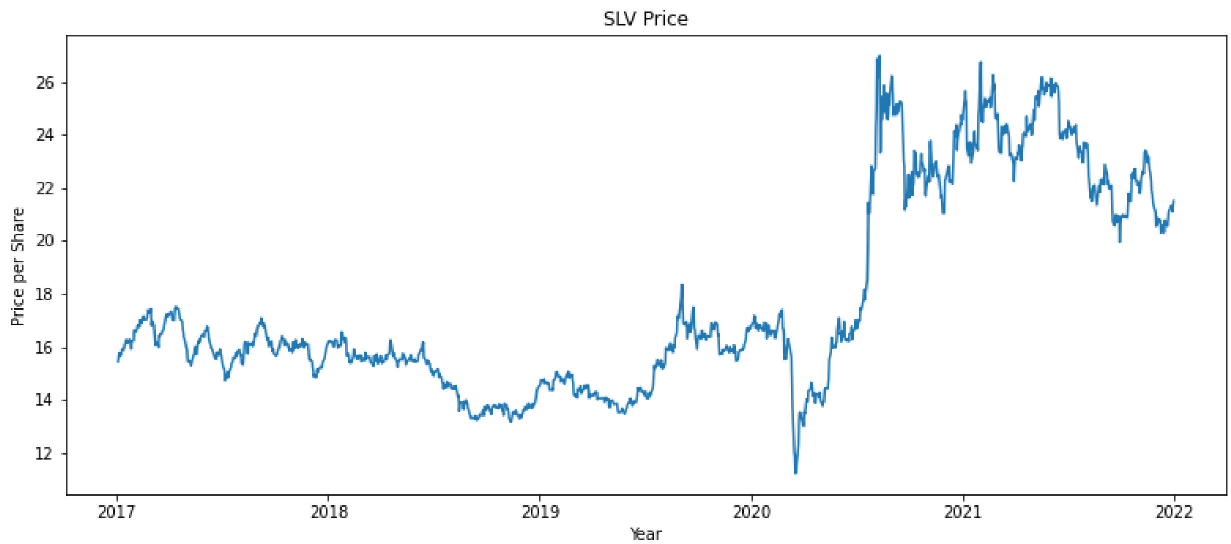

Figure 1. The SLV price chart over the time period of the study.

In Figure 2, the performance of SLV is compared with those of GLD, an ETF that tracks the price of gold, and SPY, an ETF representing the broad large-capitalization stock market since the beginning of the study period. This figure shows that SLV and GLD

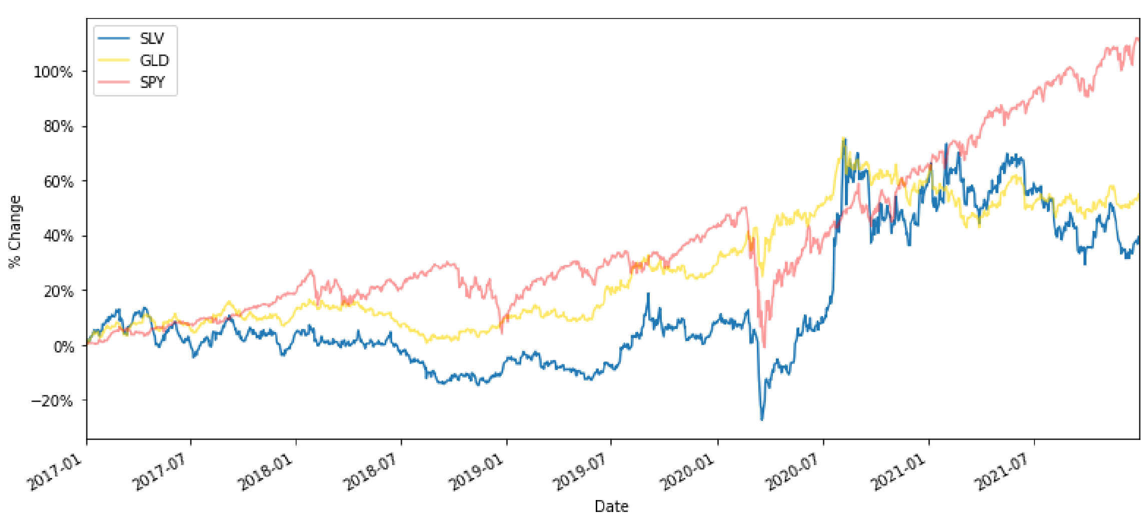

Figure 2. The percent change of SLV (blue), GLD (yellow), and SPY (pink) from the baseline value at the start of the study.

demonstrate similar patterns but with SLV exhibiting much greater volatility. Over the timeframe of the dataset, SLV has generally been outperformed by GLD and SPY but has exhibited periods of significant appreciation. Over the past year, there has been a large divergence from the performance of SPY and those of GLD and SLV, with the broad market greatly outperforming.

Stationarity

A cursory glance at the SLV chart (Figure 1) reveals that the SLV price action is not stationary. To obtain a stationary dataset, the SLV price can be differenced by the lag-1 values, producing the daily differenced dataset. The resulting plot is shown in Figure 3. To confirm the stationarity of the transformed data, the Augmented Dickey-Fuller (ADF) test is applied. The p-value of the ADF test, equal to 3.5×10^{-13} indicates that the null hypothesis that a unit root is present in the data sample can be strongly rejected, and the data are stationary. Furthermore, the very low p-value shows that a second differencing is unnecessary. The application of the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, which tests for stationarity about a deterministic trend, gives a KPSS p-value > 0.1, confirming the stationarity of the dataset.

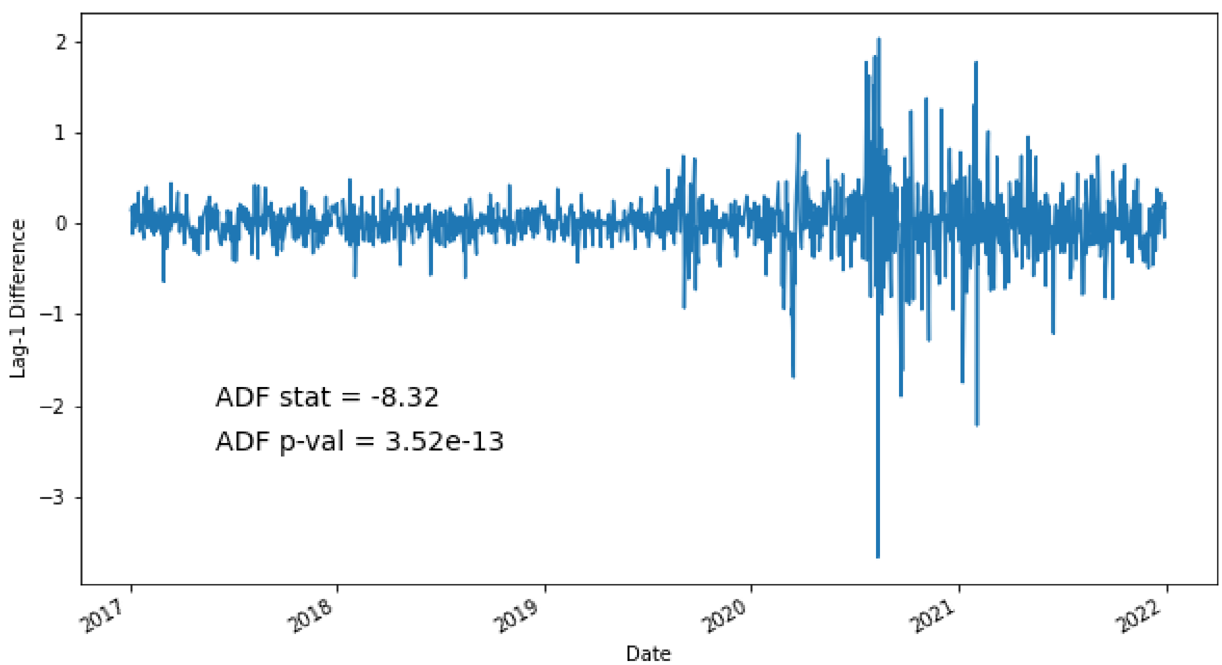

Figure 3. The lag-1 difference residual for the SLV daily price data over the time period of the study.

Autocorrelation

To get a better sense of the datasets and gain insight into the types of models needed to describe the data, autocorrelation function (ACF) and partial autocorrelation function (PACF) plots are presented, shown in Figure 4. The ACF plot shows significant autocorrelations at lags 4, 9, 15, and 16. Considering that there are five trading days in most weeks, this perhaps suggests a weekly periodicity. Furthermore, the ACF and PACF plots show very similar results. As neither the ACF nor the PACF plot displays conventional behavior indicative of an AR (autoregressive) or MA (moving average)

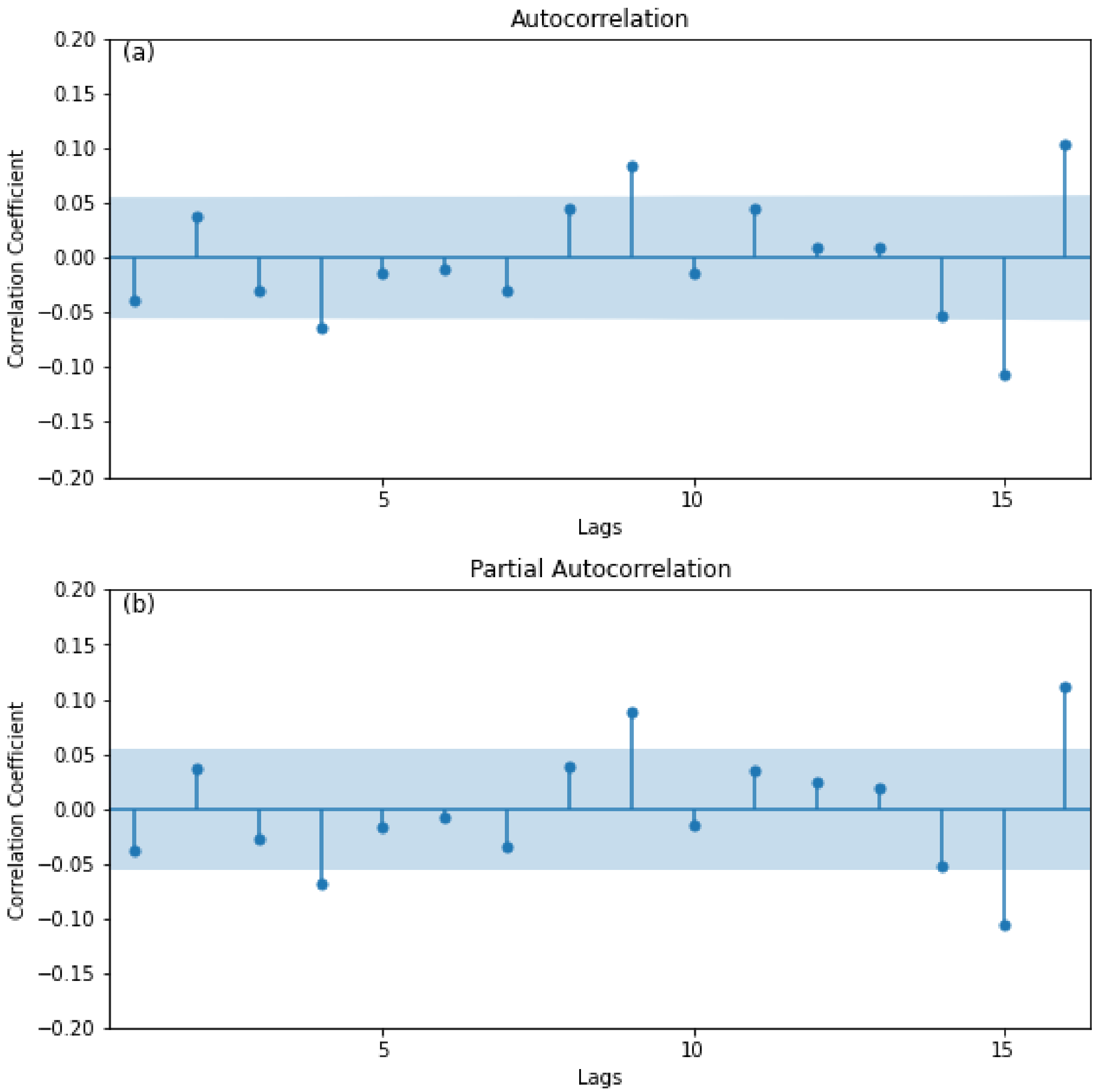

Figure 4. (a) Autocorrelation function (ACF) and (b) partial autocorrelation function (PACF) plots for the once-differenced SLV price data.

model, it is apparent that an ARIMA model of mixed components would be necessary to characterize the SLV price profile, requiring a grid search of orders for the time series model.

Modeling

Train–Test Split

To use historical SLV data in forecasting future silver prices, the SLV data is first divided into training and test sets. In this study, an 80/20 train/test ratio is used, in which the first 80% of the dataset is used as the training set upon which model parameters are determined, and the last 20% of the dataset is used as the test set upon which predictions are evaluated. Lastly, an out-of-sample forecast is generated for the time period succeeding that of the dataset.

Model-Order Fitting

To determine the requisite model parameters, an autoregressive integrated moving average (ARIMA) model is fitted onto the training set. In this process, ARIMA models of different autoregression, difference, and moving average orders are fitted, and the resulting Akaike Information Criterion (AIC) is tabulated. The model with the lowest AIC value for a given dataset is determined to be the best model for forecasting from that dataset. The Bayesian Information Criterion (BIC), which gives a greater penalty than the AIC for complex models, is used to place the models in greater context and guard against overcomplexity.

Table 3 shows the model orders for the lowest AIC models. All best-fit models require $d \geq 1$, as demonstrated in the Stationarity section and Figure 3. As the table indicates, first-order differencing is sufficient for the SLV dataset. The best AIC model is an ARIMA model described by five autoregression terms (p) and four moving average terms (q). This ARIMA(5,1,4) model characterization is not terribly surprising, considering that the first significant lag corresponds to lag 5 in the ACF and PACF plots (Figure 4). However, the ARIMA(3,1,2) model with autoregressive terms corresponding to the first three lags and moving average terms for the first two lags shows only a

	p	d	q	AIC	BIC
39	5	1	4	530.581666	579.719040
23	3	1	2	531.754321	561.236745
17	2	1	3	532.464364	561.946789
34	4	1	6	532.542914	586.594025
40	5	1	5	532.821356	586.872467

Table 3. ARIMA model orders and Akaike (AIC) and Bayesian (BIC) Information Criterion metrics for ARIMA models fit onto the training set, sorted by AIC value.

slightly higher AIC value. When the models are sorted by the BIC metric, the model corresponding to white noise with only first-differencing (ARIMA(0,1,0)) is shown to have the lowest value, as demonstrated in Table 4, although the applicability of this model to forecasting is rather limited. However, the next model in BIC value is ARIMA(3,1,2), which demonstrates that the model fits the data well and exhibits satisfactory simplicity. Note that ARIMA(5,1,4) does not appear in the top five models with respect to BIC.

	p	d	q	AIC	BIC
0	0	1	0	555.055120	559.968857
23	3	1	2	531.754321	561.236745
17	2	1	3	532.464364	561.946789
7	1	1	0	556.850132	566.677607
1	0	1	1	556.869741	566.697215

Table 4. Results of grid search of ARIMA models applied to the training set sorted by BIC value.

Heteroscedasticity

Another issue to consider is heteroscedasticity. In general, financial data tend to be quite heteroscedastic, showing periods of high volatility interspersed with periods of much greater dormancy. The SLV dataset is no exception to this, as all fitted models exhibit high levels of kurtosis, the statistical manifestation of heteroscedasticity. One way to treat this is to model the volatility through, for example, a Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model, coupled with an ARIMA model through the residuals of the ARIMA model. Another way of reducing high levels of kurtosis is to apply a log transformation to the dataset. Fitting an ARIMA model to the log(SLV price) of the daily dataset yields the best-fit models shown in Table 5. Again, ARIMA(5,1,4) does not appear in the top-five list,

	p	d	q	AIC	BIC
15	2	1	3	-5390.343511	-5360.861087
20	3	1	2	-5390.251158	-5360.768734
16	2	1	4	-5388.267291	-5353.871130
21	3	1	3	-5387.999458	-5353.603296
17	2	1	5	-5386.383013	-5347.073114

Table 5. Results of grid search of ARIMA models applied to log-transformed training dataset, sorted by AIC value.

whereas, once again, ARIMA(2,1,3) and ARIMA(3,1,2) top the list. Through the log transformation, the kurtosis for ARIMA(3,1,2) is reduced from 26.48 to 15.30, allowing for a better fit of the extreme values. The model coefficients and statistics for the model are shown in Table 6. This log model is evaluated against the models discussed above in the Prediction & Forecast section.

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                  Close   No. Observations:                 1007
Model:                 ARIMA(3, 1, 2)   Log Likelihood                2701.126
Date:                Mon, 10 Jan 2022   AIC                          -5390.251
Time:                        18:22:15   BIC                          -5360.769
Sample:                             0   HQIC                         -5379.049
                                - 1007
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.1968      0.090     13.305      0.000       1.021       1.373
ar.L2         -0.6033      0.103     -5.836      0.000      -0.806      -0.401
ar.L3         -0.0769      0.021     -3.614      0.000      -0.119      -0.035
ma.L1         -1.1601      0.084    -13.773      0.000      -1.325      -0.995
ma.L2          0.6598      0.089      7.397      0.000       0.485       0.835
sigma2         0.0003   4.96e-06     54.972      0.000       0.000       0.000
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):              6447.60
Prob(Q):                              0.98   Prob(JB):                         0.00
Heteroskedasticity (H):               5.41   Skew:                            -0.78
Prob(H) (two-sided):                  0.00   Kurtosis:                        15.30
===================================================================================
```

Table 6. Summary of the ARIMA(3,1,2) model fitted to log-transformed data.

Seasonality

As suggested by the autocorrelation plots in Figure 4, SLV appears to exhibit some seasonality in its data, on the order of about a week. To explore this possibility, Seasonal ARIMA (SARIMA) models are applied through a grid search for the optimal model parameters. Table 7 shows the results of this grid search. One seasonal model,

	p	d	q	P	D	Q	s	AIC	BIC
126	0	1	0	2	0	2	3	-5397.588665	-5373.019978
717	2	1	3	0	0	0	5	-5390.343511	-5360.861087
461	2	1	3	0	0	0	4	-5390.343511	-5360.861087
11	2	1	3	0	0	0	0	-5390.343511	-5360.861087
268	2	1	3	0	0	0	3	-5390.343511	-5360.861087

Table 7. Results of grid search of SARIMA models applied to log-transformed dataset.

of periodicity (s) equal to 3, has a lower AIC, as well as BIC score, than the non-seasonal ARIMA models. In this model, SARIMA(0,1,0)(2,0,2)3, all of the dependence on previous lags is periodic, yielding autoregressive and moving average terms representing lags 3 and 6, as shown in Table 8.

```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                                 Close   No. Observations:              1007
Model:             ARIMA(0, 1, 0)x(2, 0, [1, 2], 3)   Log Likelihood             2703.794
Date:                               Mon, 10 Jan 2022   AIC                       -5397.589
Time:                                       07:26:39   BIC                       -5373.020
Sample:                                            0   HQIC                      -5388.254
                                              - 1007
Covariance Type:                                 opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.S.L3       -1.0081      0.015    -69.397      0.000      -1.037      -0.980
ar.S.L6       -0.9748      0.016    -61.977      0.000      -1.006      -0.944
ma.S.L3        1.0380      0.016     64.222      0.000       1.006       1.070
ma.S.L6        0.9634      0.019     50.706      0.000       0.926       1.001
sigma2         0.0003   4.68e-06     58.217      0.000       0.000       0.000
===================================================================================
Ljung-Box (L1) (Q):                   3.15   Jarque-Bera (JB):              8267.02
Prob(Q):                              0.08   Prob(JB):                         0.00
Heteroskedasticity (H):               5.44   Skew:                            -1.04
Prob(H) (two-sided):                  0.00   Kurtosis:                        16.89
===================================================================================
```

Table 8. Summary of the SARIMA(0,1,0)(2,0,2)3 model.

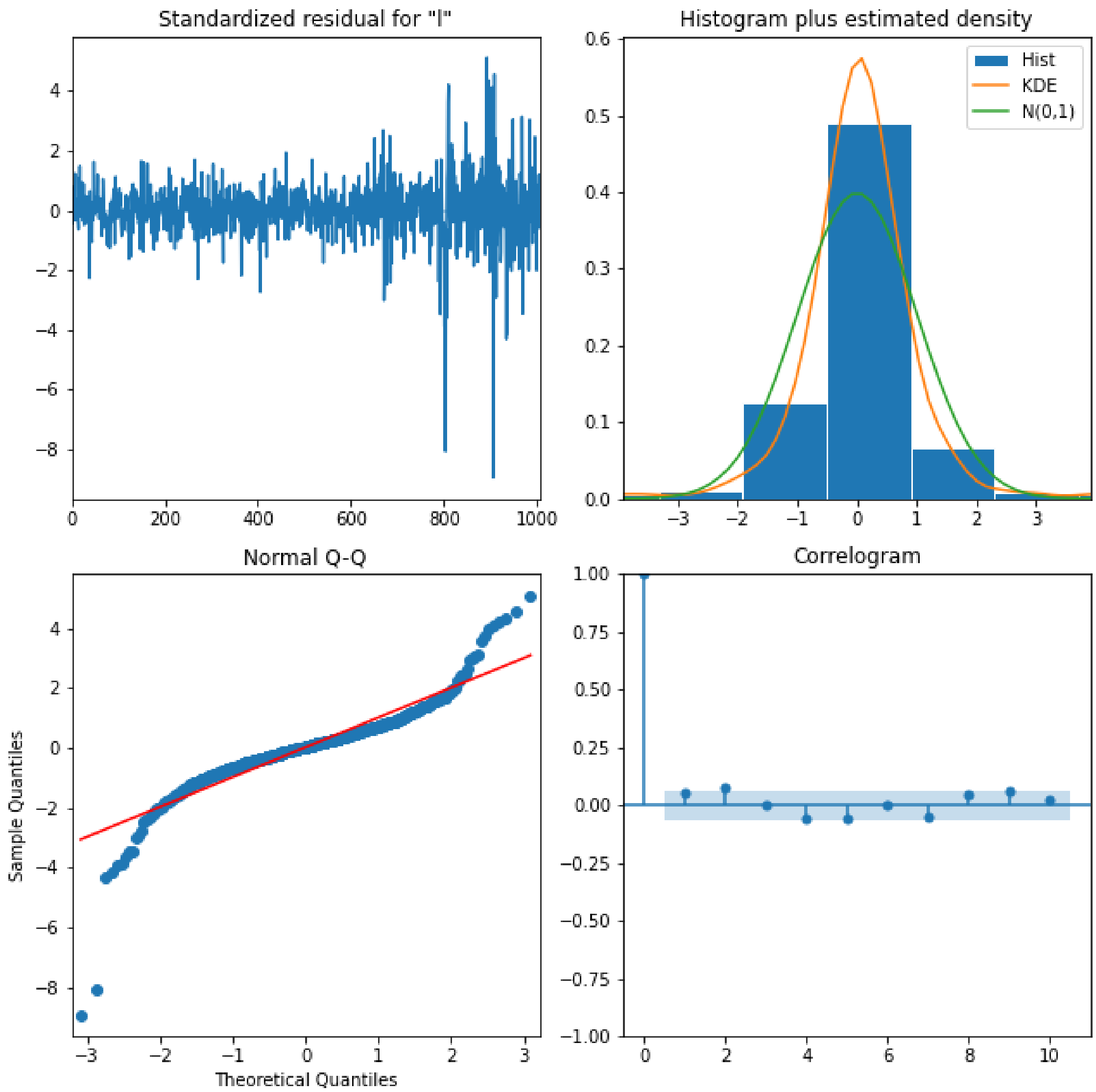

Figure 5. Diagnostic plots for the SARIMA(0,1,0)(2,0,2)3 model.

The SARIMA(0,1,0)(2,0,2)3 model is adopted as the seasonal model, with its diagnostic plots for the log-transformed version shown in Figure 5. Here, the volatility in the standardized residual plot is relatively uniform until toward the end of the training time series, where it increases considerably. The KDE plot still shows some kurtosis, exhibiting a higher peak than the normal distribution and slightly fat tails on both ends, along with minimal skew. This kurtosis is reflected in the quantile-quantile (Q-Q) plot, with extreme values veering from the normal line. Finally, the correlogram shows that the residuals do not demonstrate significant autocorrelation. This set of plots is very similar to that of ARIMA(3,1,2).

Prediction & Forecast

To determine the optimal model for forecasting the SLV performance for 2022, the selected ARIMA and SARIMA models, both in their base and log-transformed forms, are trained and evaluated with respect to the test set. The test set covers the year 2021, in which SLV suffered a down year with a performance of -12.5%. To evaluate the performance of the models, metrics are calculated on the training and test sets, using one-step-ahead predictions for the test set period. One-step-ahead predictions are generated by using the training data to make a prediction on the first test data point, adding the actual result to the training set, and training again with the new training set to generate a prediction for the next data point. Mean absolute error (MAE) and root-mean-squared error (RMSE) are the metrics applied in this study. To serve as a point of comparison, I also calculated a Baseline Model. This model predicts the exact same price as the previous day, and so can provide context for how well our other models are actually doing beyond such a simple approach.

The model results are shown in Table 9. Notwithstanding the use of different lags, the models show similar results; nevertheless, upon close examination, some conclusions can be drawn. The log-transformed models display slightly lower MAE

Model	Training Set	Test Set
Base ARIMA(3,1,2)	MAE: 0.1848 RMSE: 0.3132	MAE: 0.2997 RMSE: 0.4126
Log-transformed ARIMA(3,1,2)	MAE: 0.1843 RMSE: 0.3155	MAE: 0.3000 RMSE: 0.4157
Base SARIMA(0,1,0)(2,0,2)3	MAE: 0.1854 RMSE: 0.3110	MAE: 0.2886 RMSE: 0.4041
Log-transformed SARIMA(0,1,0)(2,0,2)3	MAE: 0.1835 RMSE: 0.3162	MAE: 0.2914 RMSE: 0.4057
Baseline	MAE: 0.1830 RMSE: 0.3184	MAE: 0.2895 RMSE: 0.4054

Table 9. Model performance results.

scores but are trained with significantly higher RMSE than their base counterparts, despite their aim in improving handling of outlier moves. This pattern continues with the evaluation test set, including in the case of the log-transformed SARIMA model, which exhibits worse MAE and RMSE scores than the base SARIMA. Only the Base SARIMA is able to train with a lower error (RMSE) than that of the Baseline, and overfitting is not

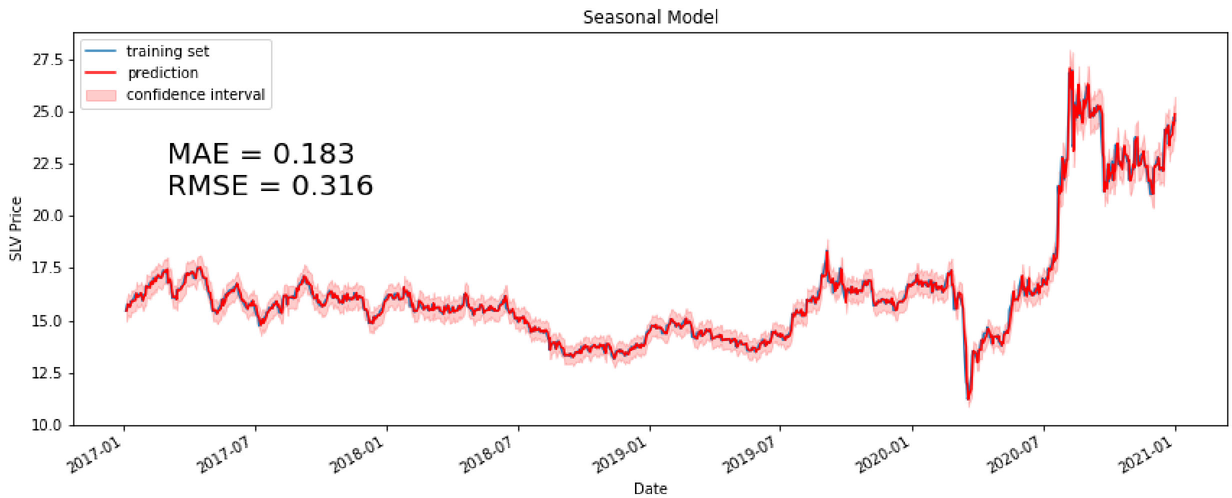

Figure 6. Model results on the training data for SARIMA(0,1,0)(2,0,2)3.

present, as this model bests the Baseline in both MAE and RMSE in the test set. The trained SARIMA results along with the corresponding 95% confidence interval are shown in Figure 6. The test set predictions for the SARIMA model and the confidence interval are displayed in Figure 7. The SARIMA model slightly outperforms the Baseline model in both MAE and RMSE for the test set. To determine whether this outperformance is statistically significant, a t-test was performed comparing the residuals of the Baseline with those of the SARIMA model with a null hypothesis asserting that the two sets of residuals are equal. A p-value of 0.86 shows that the null hypothesis cannot be rejected and that the SARIMA residuals are not statistically significantly different from the Baseline residuals. Nevertheless, the results demonstrated here indicate that the base SARIMA(0,1,0)(2,0,2)3 model is the best choice to model the SLV price data and forecast performance for 2022.

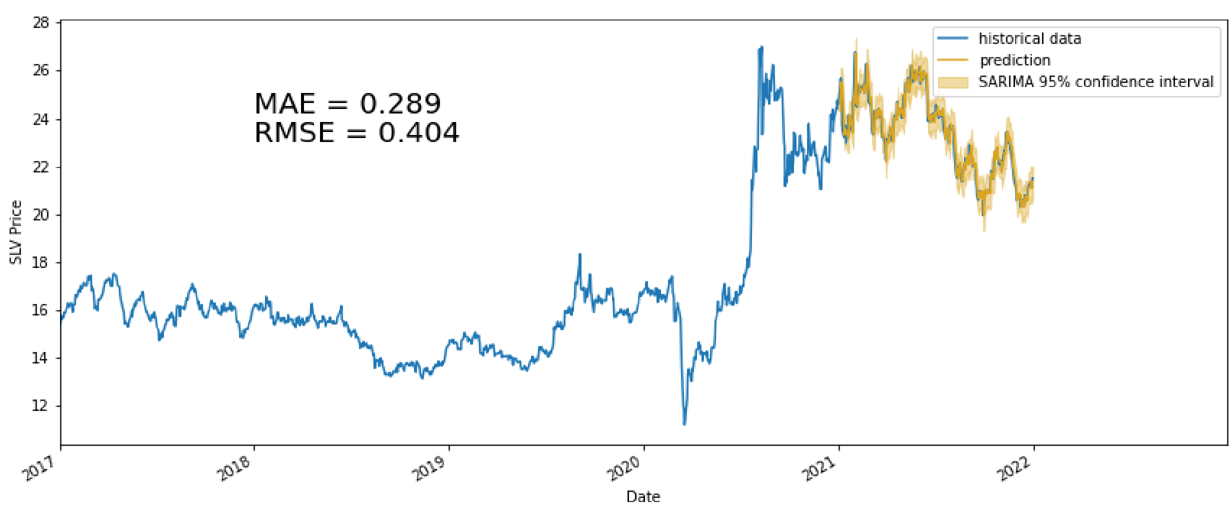

Figure 7. One-step ahead predictions of the test data from SARIMA(0,1,0)(2,0,2)3.

Looking Ahead: Forecasting 2022

Considering the forecasted outlook for 2022, we first take a look at the would-be forecast for 2021, given the training data. In this study, forecasts are generated by training with previous data to produce a forecast for a data point, adding that forecasted value to the training set, and using the new training set to generate a forecast for the next data point. Figure 8 shows the 2021 forecast along with the test set. Training with the data from 2017–2020, the SARIMA model generates a forecast of 9.6% for 2021.

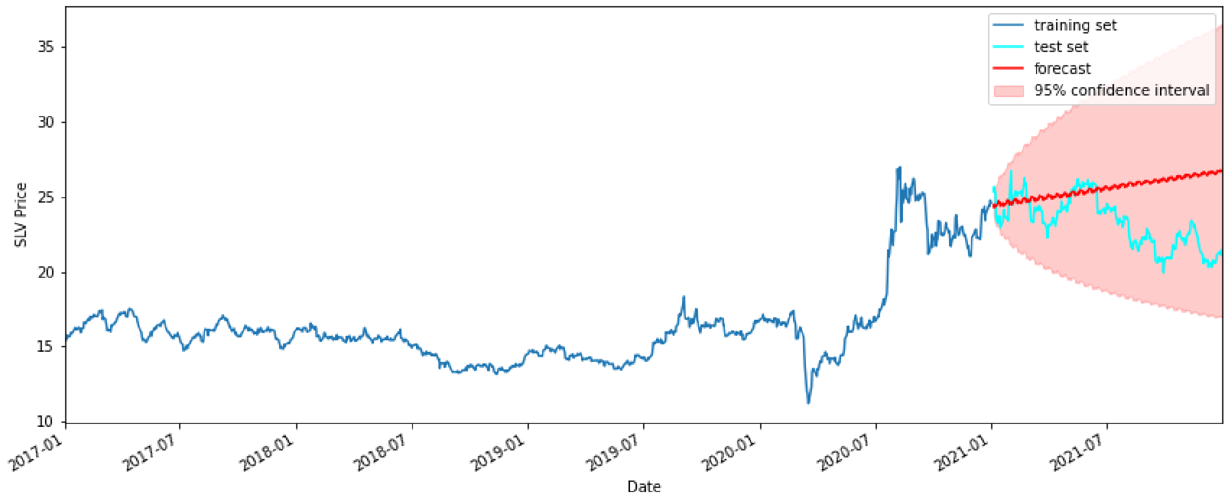

Figure 8. SLV training and test set, covering the 2021 data, along with the SARIMA forecast for 2021.

This forecast matches the SLV performance well through the first six months of 2021. However, SLV took a slide in the second half of 2021, finishing down 12.9% for the year, in contrast to the SARIMA forecast. Nevertheless, the price action throughout the year

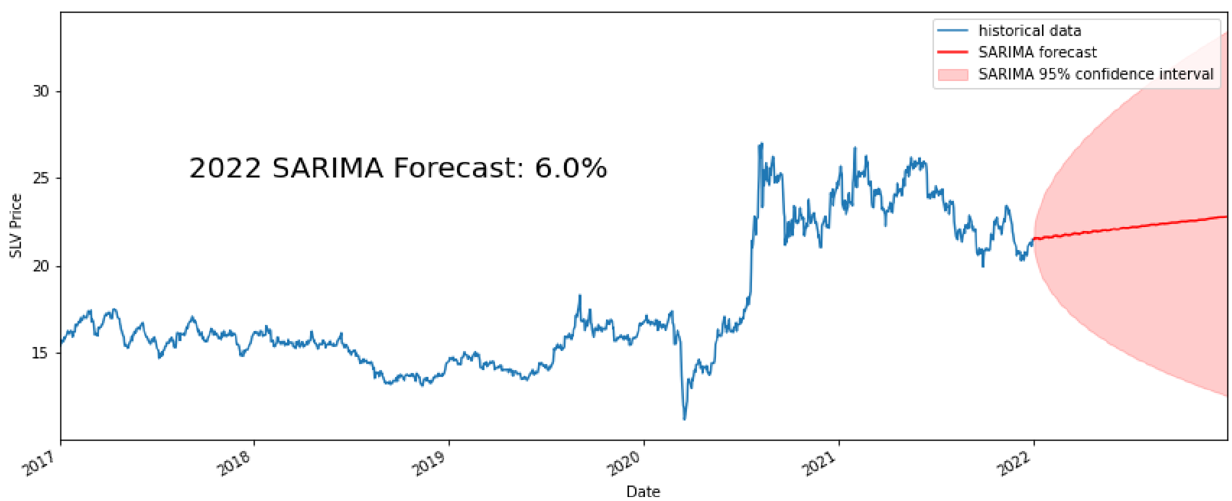

Figure 9. SLV chart and 1-year forecast for 2022 from the SARIMA model.

resides within the confidence interval of the model. For 2022, the SARIMA model forecasts an increase of 6.0% for SLV, as shown in Figure 9. This includes a very wide 95% confidence interval, with the projected price of SLV within the interval (12.56, 33.42) by the end of 2022.

Conclusions

In this study, the price of silver, through its proxy the SLV ETF, has undergone exploratory data analysis and modeling through an autoregressive-moving average time series model. From this study, the following conclusions have been reached:

- SLV daily price data are stationarized by lag-1 differencing
- SLV shows significant autocorrelation at lags 4, 9, and 15, implying a weekly seasonality
- SLV is best described by an ARIMA(3,1,2) or (2,1,3) model or a SARIMA(0,1,0)(2,0,2)3 model
- SLV data are characterized by significant heteroscedasticity
- Performing a log-transformation can reduce the heteroscedasticity but does not result in improved model performance
- SARIMA model provides the best model metrics, although not statistically significantly better than the Baseline model—a simplified model assuming tomorrow's price equals today's price
- Optimal SARIMA model forecasts an SLV appreciation of 6.0% for 2022

The results of this study, based on our optimal model, would benefit from an improved model. Such improvement could be gained by a more direct treatment of the heteroscedasticity inherent in the SLV data, perhaps by coupling the SARIMA model with a GARCH model that would model the volatility of the price data. Nevertheless, this study concludes that silver would be a good investment for 2022.