Article

A Linear Process Approach to Short-term Trading Using the VIX Index as a Sentiment Indicator

Yawo Mamoua Kobara [1,‡], Cemre Pehlivanoglu [2,‡*] and Okechukwu Joshua Okigbo [3,‡]

1 Western University; ykobara@uwo.ca
2 Cidel Financial Services; cemrepeh@gmail.com
3 WorldQuant University; okigbookey@gmail.com
* Correspondence: cemrepeh@gmail.com
‡ These authors contributed equally to this work.

Abstract: One of the key challenges of stock trading is the stock prices follow a random walk process, which is a special case of a stochastic process, and are highly sensitive to new information. A random walk process is difficult to predict in the short-term. Many linear process models that are being used to predict financial time series are structural models that provide an important decision boundary, albeit not adequately considering the correlation or causal effect of market sentiment on stock prices. This research seeks to increase the predictive capability of linear process models using the SPDR S&P 500 ETF (SPY) and the CBOE Volatility (VIX) Index as a proxy for market sentiment. Three econometric models are considered to forecast SPY prices: (i) Auto Regressive Integrated Moving Average (ARIMA), (ii) Generalized Auto Regressive Conditional Heteroskedasticity (GARCH), and (iii) Vector Autoregression (VAR). These models are integrated into two technical indicators, Bollinger Bands and Moving Average Convergence Divergence (MACD), focusing on forecast performance. The profitability of various algorithmic trading strategies are compared based on a combination of these two indicators. This research finds that linear process models that incorporate the VIX Index do not improve the performance of algorithmic trading strategies.

Keywords: Short-term trading, mean reversion, VIX, SPY, linear stochastic process, MACD, Bollinger Bands

1. Introduction

The practice of stock trading relies mainly on technical indicators that are built upon historical prices of an underlying security. As historical prices follow a random walk process, which is a special case of a stochastic process, it is difficult to make price predictions in the short-term. In order to make more accurate predictions, traders need to consider the market trends as well as the trend of an underlying security. The current market trends are heavily influenced by the advancements in technology and the subsequent increase in the use of social media. Now more than ever, short-term stock price movements are driven by a type of social anxiety called "Fear of missing out" (FOMO), fuelled by the news and social media (Khan 2021). A trading atmosphere primarily driven by emotional factors such as fear and greed reduces the efficacy and accuracy of common technical indicators built on an underlying process, thereby leading to the erosion of profits. This calls for investigating the influence of market sentiment on short-term trading.

While numerous methods exist for market sentiment analysis, such as Twitter sentiment analysis (Souza et al. 2015), relying on data collected from third parties and requiring additional modeling, the scope of this research is the CBOE Volatility (VIX) Index as a proxy for market sentiment. This paper argues that, as a leading indicator, the VIX Index improves the short-term exit and entry signal precision. A high VIX generally indicates that investors are weary of the current market environment and are purchasing insurance contracts to hedge their portfolios in anticipation of a drawdown. Whereas, a low VIX suggests complacency and that any movement in the stock market will be restrained.

This paper seeks to answer the question of whether the incorporation of the VIX index improves the performance of a mean reversion trading strategy. It proposes a two-step approach: the first step is to fit a linear process model with the multivariate time series including the VIX Index; and the second step is to integrate the linear process model into day trading with technical indicators such as Bollinger Bands and Moving Average Convergence Divergence (MACD). It uses intraday data from Refinitiv DataScope and works with the 15-second dataset of the SPDR S&P 500 (SPY) ETF (including the Open, High, Low, Close prices, as well as the Volume) and the VIX Index between January 1, 2019 and December 31, 2020.

2. THEORETICAL FRAMEWORK

The literature surrounding short-term trading and market sentiment is immense and varied. This section provides a literature review of the following: (i) the relationship between market sentiment and trading signals; (ii) the mean-reversion methods for short-term trading; and (iii) Bollinger Bands.

2.1. Market Sentiment

Sentiment analysis is a growing area of interest for financial analysts and investors. There is an increasing body of literature in behavioral finance that investigates the impact of sentiment on retail and institutional investors as well as financial market dynamics. Kearney and Liu (2014) identifies two types of sentiment: (i) market sentiment and (ii) text-based sentiment or textual sentiment. Market sentiment is the "beliefs about future cash flows and investment risks that are not justified by the facts at hand (Kearney and Liu 2014; Baker and Wurgler 2007) but based on "the subjective judgments and behavioral characteristics of investors" (Kearney and Liu 2014). Text-based sentiment is the "the degree of positivity or negativity in texts" (Kearney and Liu 2014). It may not have strictly positive-negative binary effects, but also includes other ones such as strong-weak and active-passive. As such, it includes a "more objective reflection of conditions within firms, institutions and markets” (Kearney and Liu 2014). In addition, Kearney and Liu (2014) and Li (2006) argue to include information received from text-based sentiments in modern econometric models.

Souza et al. (2015) use Twitter sentiment analytics in their investigation for a possible significant relationship between Twitter sentiment information and financial market dynamics - stock returns, volatility, and volume (Souza et al. 2015). Their results indicate that the social media is a valuable source of information, especially in the retail sector (Souza et al. 2015). On that account, Twitter real-time information "if properly modelled, can provide ex-ante information about the market even before the main news wires (Souza et al. 2015)."

Moreover, prominent institutions have developed tools and methods to measure and quantify market sentiment. These include social media sentiment analysis, Chicago Board Options Exchange (CBOE) Volatility Index (VIX) (Bekiros and Georgoutsos 2008; Li 2006; Corrado and Miller 2005; Fernandes et al. 2014), NYSE Bullish percentage (Clarke and Statman 1998), Baker and Wurgler sentiment index (BW index) (Baker and Wurgler 2006), Buffet Indicator Chang and Pak (2018); Mislinski (2021), etc.

On September 22, 2003, the Chicago Board Options Exchange (CBOE) introduced a new CBOE Volatility (VIX) Index, replacing the older volatility index. The CBOE VIX calculation is based on the prices of a batch of (OTM) out-of-the-money and (NTM) near-the-money put and call options on the S&P 500 index (SPY). Thus, CBOE Volatility Index (VIX) "is simply the price of a linear portfolio of options." (Li 2006) The VIX Index is given by Nielsen (Nielsen):

$$\sigma^2 = \frac{2}{T}\sum_{i}^{n}\frac{\Delta K_i}{K_i^2}e^{RT}Q(K_i) - \frac{1}{T}\left[\frac{F}{K_0} - 1\right]^2 \tag{1}$$

where: $\sigma = VIX/100 \Rightarrow VIX = \sigma \times 100$, T = expiration time, F = forward index level based on index option prices, K_0 = first strike below F, K_i = strike price of i^{th} out-of-money option; a call if $K_i > K_0$, put if $K_i < K_0$; both call and put if $K_i = K_0$, $\Delta K_i =$ interval between strike prices - half the difference between the strike on either side of k_i, such that

$$\Delta k_i = \frac{k_{i+1} - k_{i-1}}{2} \tag{2}$$

R = risk-free interest rates to expiration, and $Q(K_i)$ = the average of the bid quote and ask quote for each option with strike K_i.

Lei et al. (2012) explore the relationship between the VIX as a proxy for market sentiment and trading volume. VIX measures the expected 30-day implied volatility, and is often referred to as "investor fear gauge". The authors find empirical evidence to suggest that increases in the VIX Index explain the percentage increase in trading volume, but only during a high VIX period.

2.2. Mean Reversion

Mean reversion is a pertinent financial theory to explain how prices will evolve over time. Arguably, the most straightforward definition of mean reversion is the observation of the tendency of asset prices to fall (rise) after reaching a high (low) level over time. Lee (1991) describes mean-reverting returns as returns that are negatively autocorrelated. The discrete description of the model is given by

$$R_n = \alpha\Big(R_{(n-1)} - \mu\Big) + \mu + \sigma Z_n \tag{3}$$

where R_n is the return at time n, μ is the average return, and Z_n is a white noise with variance σ, and $\alpha < 1$ is the autocorrelation coefficient.

In statistics, a stationary distribution exhibits mean reversion. In other words, a return model is mean-reverting if the sequence of returns is a stationary process. A process is a strictly stationary time series if the time lag of the process has the same joint distributions. Mathematically (Lee 1991), the process $\{X_t\}$ is strictly stationary, if

$$F(X_1, \ldots, X_n) = F(X_{1+h}, \ldots, X_{n+h}) \tag{4}$$

for all integers h *and* $n \geq 1$. A weaker version of stationarity is obtained when the first and second moments are time independent. That is, the mean and covariance do not depend on time.

2.3. Bollinger Bands

Bollinger Bands are volatility bands that are based on a moving average. A simple moving average (SMA) is given by the following:

$$S_t = \sum_{j=1}^{\infty}\phi_j Z_{t-j} \tag{5}$$

for all t, where Z_t is a white noise with mean zero and variance σ^2, and ϕ_j is a sequence of parameters such that $\sum_{j=1}^{\infty}|\phi_j| < \infty$. The upper and lower bands are $\pm m\,\sigma$ away from the mean, where $m \in R^+$. A breakout beyond the bands are considered significant moves that may be used to predict future prices. Steven Gold (2018) explores the following price behaviors associated with Bollinger Bands: Trend-following (Momentum), Contrarian (Mean-reversion), and Squeeze. The Squeeze viewpoint argues that periods of low volatility are followed by periods of high volatility. Gold (2018) suggests that algorithmic trading strategies based on the Bollinger Bands, with and without the Squeeze Effect, yield positive returns that exceed a Buy-and-Hold investment strategy.

3. METHODOLOGY

This section provides a general overview of the following linear process models that are being used to predict the future conditional mean and volatilities of SPY prices: (i) the Autoregressive Integrated Moving Average (ARIMA); (ii) the Vector Autoregressive (VAR); and (iii) Generalized Autoregressive Conditional Heteroskedasticity (GARCH). In general, time series models are stochastic processes modeled as a linear or a nonlinear model. The Autoregressive (AR) and Moving Average (MA) models are widely used linear models, when combined provide the so-called Autoregressive Moving Average (ARMA) model or Autoregressive Integrated Moving Average (ARIMA) model after its integration. Models with trend, drift, and seasonality are grouped under the Seasonal Autoregressive Integrated Moving Average (SARIMA) models. Box-Jenkins (Box et al. 1970) have popularized these models and employed the following ARMA(p, q) model:

$$y_t - \sum_{i=1}^{p} \phi_i \, y_{t-i} - \phi_0 = \theta_0 + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon \tag{6}$$

The ARMA models are only used for stationary time series. When the dataset exhibits non-stationarity, the ARIMA(p,d,q) is appropriate. ARIMA(p,d,q) is an ARIMA(p,q) fitted after d differences on the dataset. In ARIMA models a non-stationary time series is made stationary by applying finite differencing of the data points. The $ARIMA(p,d,q) \times (P,D,Q)^s$ model is a generalization of an ARIMA model to include the case of non-stationarity, trend, drift, and seasonality.

As a standard multivariate data analysis tool in econometrics, the VAR model has evolved from the basic univariate linear AR process to the structured VAR and Vector Error Correction (VEC) models over the years (Amisano and Giannini 2012; Lütkepohl 2005). In its simplest form, it consists of a set of N time series $y_t = (y_{1t}, \cdots, y_{nt})$ for $n = 1, \cdots,$ N. The $VAR(p)$ model is defined as:

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} y_{kt} + u_t \tag{7}$$

where A_i are $(N \times N)$ matrices for $i = 1, \cdots, p$ and u_t is an N-dimensional white noise process with $E(u_t) = 0$ and covariance matrix $E\left(u_t u_t^T\right) = \Sigma_u$. Similar to the AR model, the VAR model is a linear combination of the past observations of all the time series. However, unlike the AR model, it considers predictive causality among variables. This useful property of the VAR helps with modeling the dynamic relationships in the dataset.

In order to determine if there is a statistically significant relationship between variables, the Granger Causality test points out if a variable has predictive capability over another variable or a group of variables. It sets forth a mathematical hypothesis of a test on a dataset (Granger 1980). The hypothesis posits that the series X does not cause the series Y if

$$P(Y_{t+1} \in A / I(t)) = P(Y_{t+1} \in A / I_{-X}(t)) \tag{8}$$

for all $t \in Z$; otherwise, the series X is said to Granger cause the series Y. $P(X)$ is the probability of event X, A is any given non-empty set, $I(t)$ is the set of all information known at time t, and $I_{-X}(t)$ is the set of information available at time t without the information on the series $X(t)$. The null hypothesis of the Granger test is that the estimated parameters of models are equal to zero. In reality, Granger causality does not imply true causality. It is only a mathematical predictive property based on the principles that causality happens before its effect and the past has unique information about the future of the effect.

Followed by the Granger causality test, the cointegration test is considered to determine the existence of a statistically significant correlation between the variables. In mathematical terms, there is a cointegration between two stationary time series, X_t and Y_t, if the linear combination of the series can be described as a stationary process:

$$U_t = Y_t - \alpha X_t \tag{9}$$

If time series variables are cointegrated, then a VEC model is appropriate for modelling; otherwise, a VAR model should be used. Since the use of a stationary time series has been proven more effective in forecasting, the Johansen test can be used to test the stationarity as well as the existence of cointegration among the time series. The null hypothesis of this test is H_0: no cointegration exists, with the alternative H_1: cointegration exists. This can be done with a t-test on the coefficient α and a stationary test on U_t if α is significant.

In time-series, non-stationary processes have joint probability distributions that change over time. A non-stationary dataset whose mean and variance change over time violate the assumption of mean-reversion and render predictions difficult. To test for non-stationary in the dataset, the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test is performed, with the null hypothesis that a time series is stationary around a deterministic trend.

Linear models, whether univariate or multivariate, do not consider nonlinearity in the time series. There are important models used to model nonlinear means such as the ARCH model stipulated as (Handout and Perrelli Handout and Perrelli):

$$y_t = \sqrt{w + \sum_{i=1}^{p} \alpha_i \sigma_{t-i}^2 e_{t-i}^2} \tag{10}$$

where $r_t = \sigma_t e_t$ is the return at time t, e_t is a white noise with zero mean and variance of one, and w is a constant. This heteroskedastic nonlinear model in variance has several variations, like GARCH, EGARCH, etc. The volatility of a linear combination may become large, thus reducing the sample of possible combinations (deviations from the mean) when constructing the Bollinger Bands.

This research also considers the volatility of stock prices for use in algorithmic trading. ARIMA models the conditional expectation of a time series, but not the conditional variance. The conditional variance of ARIMA models is a constant. GARCH models provide a stochastic volatility that can be used to model the conditional variance of a stochastic process. GARCH models the conditional variance structurally as the ARIMA model of the conditional expectation.

The GARCH model was introduced by Robert Engle (1982)(Bollerslev 1986). The model is as follows:

$$y_t = e_t \sigma_t \tag{11}$$

where e_t is assumed to be a white noise (i.i.d. random variable with expectation 0 and variance 1) and are assumed independent from σ_t and

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \cdots + \alpha_p y_{t-p}^2 \tag{12}$$

Tim Bollerslev (1986) extended the ARCH(p) model to the GARCH(p,q) by setting σ_t as:

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \cdots + \alpha_p y_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_q \sigma_{t-p}^2 \tag{13}$$

In order to assess whether the models generalize well, the dataset is divided into three: training, validation, and test. First, the training dataset is used to estimate the model parameters. Second, the validation dataset is used to validate the predictive capability of the model. Lastly, the test dataset is used to determine algorithmic trading strategies. The "best" models to predict the SPY are selected based on the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC), with the expectation that the model with the lowest AIC or BIC performs the best out-of-sample. The mean absolute error (MAE), the mean squared error (MSE), and the root mean squared error (RMSE) are used as the criteria for measuring the forecast performance of the models fitted on the validation dataset. The MAE is obtained as:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|e_t| \tag{14}$$

The MSE is obtained as:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}e_t^2 \tag{15}$$

The RMSE is obtained as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}e_t^2} \tag{16}$$

where $e_t = y_t - \hat{y}_t$ is the deviation of the forecasted values from the actual ones, n is the number of iterations, y_t is the actual value, and $\hat{y}_t$ is the forecast value.

After explaining the linear process models, this section presents the implementation and the results of the models that perform best out-of-sample. These models are used to predict SPY mean and variance for the implementation of trading strategies. The models are run and the plots are generated using the Python programming language and relevant libraries (cypowered 2021).

3.1. Preliminary Results

Table 1 presents the descriptive statistics of the entire dataset. The dataset is a 2-year multivariate time series (617,360 observations). The average SPY price is around 306, and the average VIX price 23.77. Figure 2 shows that there is a weak linear relationship between SPY Last and VIX Last. The individual values are not normally distributed. The time series plots are provided in Figure 1, which shows an increasing trend in both the SPY and VIX prices. Moreover, it suggests that SPY and VIX prices are inversely related - as the value of VIX increases, SPY decreases respectively. The initial exploration of the dataset shows that VIX Granger causes SPY at lag 3. The Granger Causality at other lags is tabulated in the Table 4.

The level, first-differenced, and second-differenced time series of SPY and VIX Last prices are tested for stationarity. The results are shown in the Table 3. The p-value of the KPSS test that the SPY and VIX Last vectors are stationary is 0.01, so there is enough evidence to reject the null hypothesis in favor of the alternative of a unit root at the 95% confidence level. The p-value of the KPSS test that the first- and second-differenced SPY and VIX Last vectors are stationary is 0.1. Therefore, there is not sufficient evidence to reject the null hypothesis that the first- and second-differenced datasets are stationary. In conclusion, the first- and second-differenced datasets are stationary, and are appropriate for fitting time series models.

Based on the Granger Causality Test on first-differenced dataset, lag 2 is statistically significant, and at lag 3 the F-statistic gets smaller, before getting larger again at lag 4. Based on this a VAR(2) model appears to be more parsimonious and complete. However, a VAR(1) model fitted on the second-differenced dataset yields the best results with statistically significant coefficients. In addition to VAR, ARIMA(2,1,1) and GARCH(1,1) models are fitted, and the results are presented in the next three sections.

Table 1: Descriptive Statistics

Stats	SPY Open	SPY High	SPY Low	SPY Last	Volume	Close Bid	Close Ask	VIX Last
count	617360	617360	617360	617360	617356	614974	614974	617360
mean	305.72	305.72	305.71	305.72	3746.12	305.74	305.75	23.77
std	30.01	30.01	30.01	30.01	23592.33	30.03	30.03	11.98
min	218.45	218.46	218.31	218.32	21.00	218.31	218.33	11.44
25%	285.32	285.33	285.31	285.32	400.00	285.32	285.33	15.02
50%	300.30	300.31	300.30	300.30	1200.00	300.31	300.32	20.86
75%	327.51	327.52	327.51	327.52	3278.00	327.57	327.58	28.12
max	374.61	374.66	374.58	374.63	8951774.00	374.62	374.64	85.47

Table 2: Correlation Matrix for the First-Differenced Dataset

	SPY Last	VIX Last
SPY Last	1.0000	-0.4351
VIX Last	-0.4351	1.0000

Table 3: KPSS Test for Stationarity

KPSS Test p-values	SPY Last	VIX Last
Level	0.01	0.01
First-differenced	0.10	0.10
Second-differenced	0.10	0.10

Table 4: Granger Causality Test

Lag	SPY raw data p-value	SPY 1st difference p-value
1	0.929	0.200
2	0.446	0.040
3	0.010	0.008
4	0.020	0.000
5	0.000	0.000
6	0.000	0.000

Figure 1. SPY Open price vs VIX

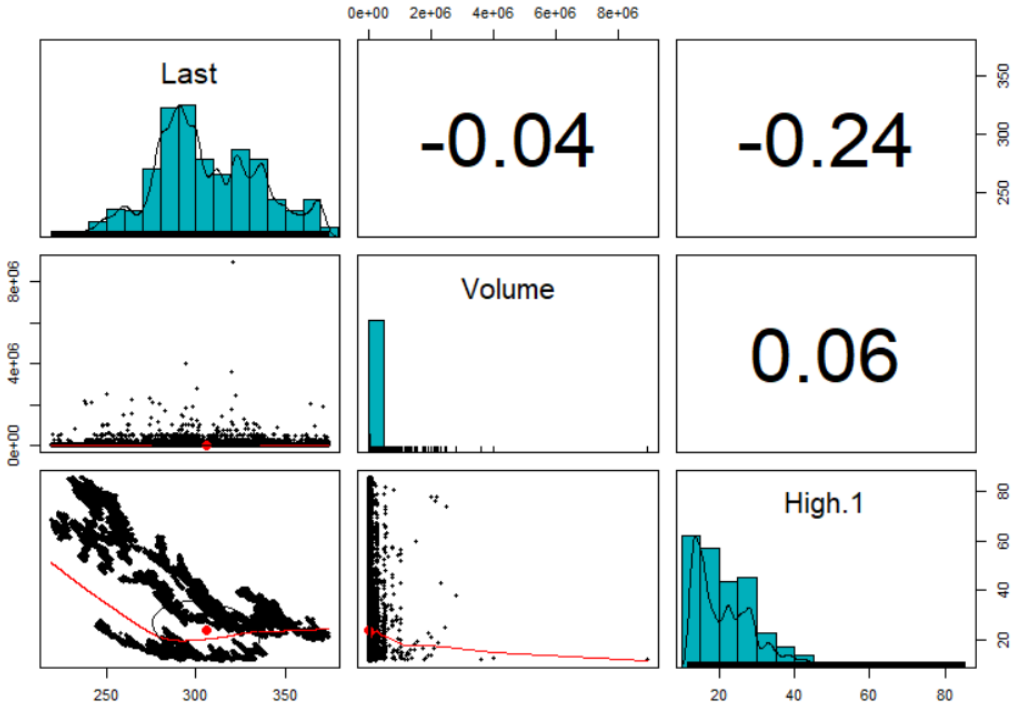
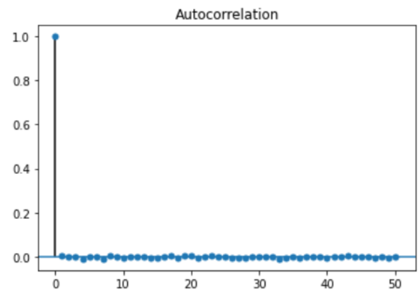
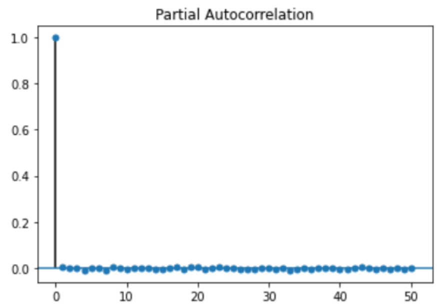


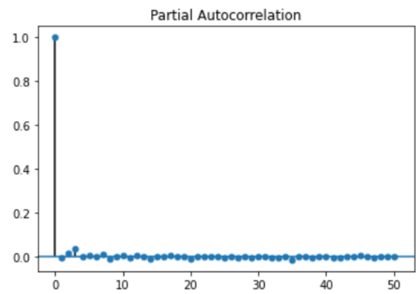
Figure 2. Pair Panel Plot of SPY Last, SPY Volume and VIX



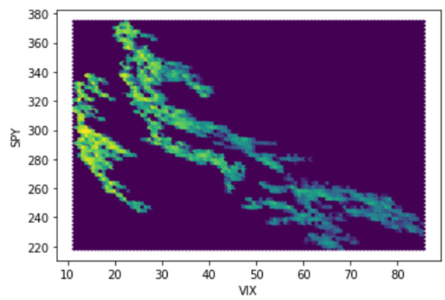
(a) ACF of SPY Open First Difference



(b) PACF of SPY Open First Difference



(c) PACF of VIX First Difference



(d) SPY vs VIX

Figure 3. SPY-VIX visualization

3.2. ARIMA Model

The first model fitted is the ARIMA model on the level dataset due to its popularity and ease. Using an auto selection algorithm based on the AIC, we find ARIMA(2,1,1) to be the best among other models vis-à-vis forecast performance. Summary results are given in the Tables 5 and 6.

Table 5: ARIMA without VIX Results

Dep. Variable:	SPY_Last	**No. Observations:**	61736
Model:	ARIMA(2, 1, 1)	**Log Likelihood**	-75663268.781
AIC	151326545.561	**BIC**	151326581.684
HQIC	151326556.767		

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
ar.L1	0.0033	6.85e-05	47.432	0.000	0.003	0.003
ar.L2	-0.0017	4.88e-05	-34.849	0.000	-0.002	-0.002
ma.L1	-1.0000	3.99e-09	-2.51e+08	0.000	-1.000	-1.000
sigma2	0.0185	1.79e-06	1.03e+04	0.000	0.018	0.018

Table 6: ARIMA with VIX Results

Dep. Variable:	SPY_Last	**No. Observations:**	61736
Model:	ARIMA(2, 1, 1)	**Log Likelihood**	52914.768
AIC	-105819.536	**BIC**	-105774.383
HQIC	-105805.530		

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
VIX_Last	-0.6930	0.001	-935.296	0.000	-0.694	-0.692
ar.L1	0.0893	0.132	0.678	0.498	-0.169	0.348
ar.L2	-0.0100	0.014	-0.710	0.478	-0.038	0.018
ma.L1	-0.1873	0.132	-1.421	0.155	-0.445	0.071
sigma2	0.0150	3.83e-06	3924.220	0.000	0.015	0.015

In the Tables 5 and 6, the coefficients of the ARIMA models are extracted and the models thus obtained are shown in equations 17 and 18 as follows:

$$\text{SPY}_t = -0.0017 \times \text{SPY}_{t-2} + 0.0033 \times \text{SPY}_{t-1} - \epsilon_{t-1} \tag{17}$$

$$\text{SPY}_t = -0.01 \times \text{SPY}_{t-2} + 0.0893 \times \text{SPY}_{t-1} - 0.693 \times \text{VIX}_t - 0.1873\epsilon_{t-1} \tag{18}$$

where SPY_t represent the SPY price at time t, VIX_t is the VIX index at time t, and ϵ_t is the residuals at time t.

In this research, Python libraries (i.e., statsmodels) are used to estimate these coefficients. In general, most software packages use the maximum likelihood estimation methods to estimate these parameters. From the output of the models tabulated in the Tables 5 and 6, the coefficients of the model without VIX are significant at the 95% confidence interval. This is confirmed by the p-values of 0. Conversely, the coefficients of the model with VIX are not significant.

Using one-step-ahead forecasting on the validation dataset, we compute the MSE, MAE and RMSE of ARIMA(2,1,1) with and without VIX to evaluate the models. Even though the model with VIX has statistically insignificant coefficients on the AR and MA variables, it performs better than its counterpart out-of-sample.

Table 7: ARIMA Model Evaluation result

	MSE	**MAE**	**RMSE**
Without VIX	47.187	5.44	6.869
With VIX	2.1512	0.0557	1.4667

3.3. VAR Model

We fit several VAR models to the second-differenced dataset. As the number of lags in a model increases, the model AIC decreases in a linear fashion. For simplicity, parsimony and to avoid over-fitting, we only fit VAR(1), VAR(2), and VAR(3) models on the second-differenced dataset. The estimated coefficients, along with the prediction accuracy metrics are given in the Tables 8, 9, and 10.

Table 8: VAR(1) model coefficients

		VAR(1) Model		
AIC:	-8.27182		N.obs:	432,149
		Results for equation SPY		
	Coefficients	**Std.Error**	**t-stat**	**p-value**
L1.SPY	-0.479653	0.001395	-343.792	0.000
L1.VIX	0.08504	0.002187	38.883	0.000
		Results for equation VIX		
	Coefficients	**Std.Error**	**t-stat**	**p-value**
L1.SPY	-0.089119	0.00087	-101.843	0.000
L1.VIX	-0.553746	0.001372	-403.685	0.000

Table 9: VAR(2) model coefficients

		VAR(2) Model		
AIC:	-8.55401		N.obs:	432,148
		Results for equation SPY		
	Coefficients	**Std.Error**	**t-stat**	**p-value**
L1.SPY	-0.651488	0.001552	-419.662	0.000
L1.VIX	0.058239	0.002475	23.530	0.000
L2.SPY	-0.312047	0.001569	-198.943	0.000
L2.VIX	0.080834	0.002452	32.964	0.000
		Results for equation VIX		
	Coefficients	**Std.Error**	**t-stat**	**p-value**
L1.SPY	-0.118064	0.000960	-122.958	0.000
L1.VIX	-0.762670	0.001531	-498.178	0.000
L2.SPY	-0.076662	0.000970	-79.020	0.000
L2.VIX	-0.392182	0.001517	-258.573	0.000

Table 10: VAR(3) model coefficients

		VAR(3) Model		
AIC:	-8.68377		N.obs:	432,147
		Results for equation SPY		
	Coefficients	**Std.Error**	**t-stat**	**p-value**
L1.SPY	-0.733171	0.001614	-454.132	0.000
L1.VIX	0.048647	0.002610	18.637	0.000
L2.SPY	-0.471951	0.001933	-244.209	0.000
L2.VIX	0.077455	0.003098	25.004	0.000
L3.SPY	-0.216067	0.001641	-131.629	0.000
L3.VIX	0.083551	0.002574	32.459	0.000
		Results for equation VIX		
	Coefficients	**Std.Error**	**t-stat**	**p-value**
L1.SPY	-0.132478	0.000997	-132.883	0.000
L1.VIX	-0.861947	0.001612	-534.760	0.000
L2.SPY	-0.115711	0.001193	-96.959	0.000
L2.VIX	-0.588190	0.001913	-307.488	0.000
L3.SPY	-0.058633	0.001014	-57.843	0.000
L3.VIX	-0.264253	0.001590	-166.249	0.000

We obtain the VAR model equations by extracting the coefficients from the Tables 8, 9, and 10. The VAR(1) model equation is obtained as follows:

$$\text{SPY}_t = -0.479653 \times \text{SPY}_{t-1} + 0.08504 \times \text{VIX}_{t-1} \tag{19}$$

The VAR(2) model equation is obtained as follows:

$$\begin{aligned} \text{SPY}_t = &-0.651488 \times \text{SPY}_{t-1} + 0.058239 \times \text{VIX}_{t-1} \\ &-0.312047 \times \text{SPY}_{t-2} + 0.080834 \times \text{VIX}_{t-2} \end{aligned} \tag{20}$$

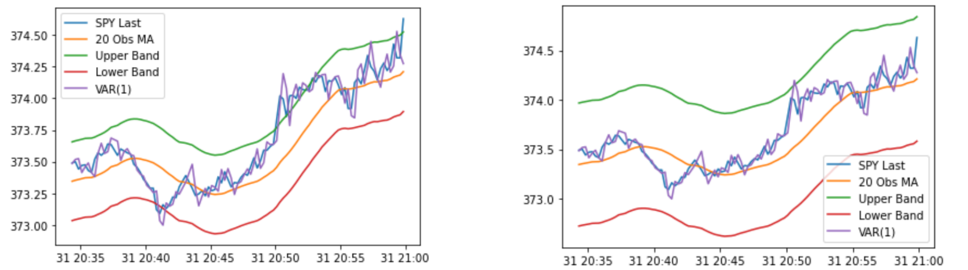
And finally the VAR(3) model equation is obtained as follows:

$$\begin{aligned} \text{SPY}_t = &-0.733171 \times \text{SPY}_{t-1} + 0.048647 \times \text{VIX}_{t-1} - 0.471951 \times \text{SPY}_{t-2} + \\ &0.077455 \times \text{VIX}_{t-2} - 0.216067 \times \text{SPY}_{t-3} + 0.083551 \times \text{VIX}_{t-3} \end{aligned} \tag{21}$$

Using one-step-ahead forecasting on the validation dataset, we compute the MSE, MAE, and RMSE of the models and provide them in the Table 11. All three measures suggest VAR(3) model performs the best out-of-sample. However, we use the VAR(1) model as our basis for forecasting to have the same number of variables in other models for comparability and simplicity. The Figures 4 (a) and (b) visualize the forecasting performance of the VAR(1) model.

Table 11: Model accuracy measures

	RMSE	MSE	MAE
VAR(1)	0.13271	0.17613	0.07507
VAR(2)	0.12521	0.01568	0.07079
VAR(3)	0.12136	0.01472	0.06840



(a) Actual vs VAR(1) with its 68% confidence interval

(b) Actual vs VAR(1) with its 95% confidence interval

Figure 4. VAR(1) Forecast Model

3.4. GARCH Model

This sub-section presents the GARCH(1,1), GARCH(2,1), and GARCH(2,2) models fitted on the SPY Last dataset. The model results are provided in the Tables 12, 13, and 14. GARCH models are used to estimate and forecast the SPY Last price volatility, since there are varied periods of high and low variance. The models could alternatively be fitted on the residuals of the ARIMA model as well and produce the same results.

Table 12: GARCH(1,1) model coefficients

		GARCH(1,1)		
AIC:	3.98 e+06		N.obs:	432,148
		Mean Model		
	Coefficients	**Std.Error**	**t-stat**	**p-value**
mu	292.4498	0.0375	7799.39	0.000
		Volatility Model		
	Coefficients	**Std.Error**	**t-stat**	**p-value**
Omega	9.85	0.0226	435.544	0.000
alpha[1]	0.1998	0.0000536	372.713	0.000
beta[1]	0.7794	0.0000525	1485.269	0.000

Table 13: GARCH(2,1) model coefficients

		GARCH(2,1)		
AIC:	3.76 e+06		N.obs:	432,148
		Mean Model		
	Coefficients	**Std.Error**	**t-stat**	**p-value**
mu	289.854	0.00145	19960	0.000
		Volatility Model		
	Coefficients	**Std.Error**	**t-stat**	**p-value**
Omega	0.000865	0.000377	2.289	0.0022
alpha[1]	0.7082	0.05166	13.708	0.000
alpha[2]	0.000	0.175	0.000	1.000
beta[1]	0.2918	0.179	1.63	0.103

For GARCH(1,1), the model equation is:

Table 14: GARCH(2,2) model coefficients

	GARCH(2,2)			
AIC:	3.76 e+06		N.obs:	432,148
	Mean Model			
	Coefficients	**Std.Error**	**t-stat**	**p-value**
mu	288.57	0.00245	19960	0.000
	Volatility Model			
	Coefficients	**Std.Error**	**t-stat**	**p-value**
Omega	0.001326	0.000283	4.688	0.0000028
alpha[1]	0.7082	0.0744	9.416	0.000
alpha[2]	0.000	0.0013	0.000	1.000
beta[1]	0.000	0.00164	0.000	1.000
beta[2]	0.2986	0.073	4.068	0.000047

$$\text{Var}(SPY_t|SPY_{t-1}) = 9.85 + 0.1998 \times SPY_{t-1}^2 + 0.7794\ \sigma_{t-1}^2 \tag{22}$$

For GARCH(2,1), the model equation is:

$$\text{Var}(SPY_t|SPY_{t-1}, SPY_{t-2}) = 0.001 + 0.708 \times SPY_{t-1}^2 + 0.292\ \sigma_{t-1}^2 \tag{23}$$

For GARCH(2,2), the model equation is:

$$\text{Var}(SPY_t|SPY_{t-1}, SPY_{t-2}) = 0.001 + 0.708 \times SPY_{t-1}^2 + 0.299\ \sigma_{t-2}^2 \tag{24}$$

In equations 23 and 24, the $\text{Var}(SPY_t|SPY_{t-1}, SPY_{t-2})$ stands for the *variance* of SPY conditional on the previous two observations and is not to be confused with the VAR model.

4. TRADING STRATEGIES

This section proposes and compares five trading strategies. We use the Buy-and-Hold investment strategy as a benchmark for all trading strategies. These can be summarized as follows:

- Simple Bollinger Bands
- Bollinger Bands + VAR
- Bollinger Bands (2 SD GARCH Volatility) + VAR
- Bollinger Bands + 1 Min MACD
- Bollinger Bands + VAR + 1 Min MACD

The trading strategies are based on, but not limited to, the following assumptions:

- Only one share traded at a time
- No shorting
- No slippage
- No dividends
- No fees or commissions
- No taxes
- No corporate actions (i.e., stock splits)
- Trades are fulfilled immediately and at the price the signals were generated

In addition, all trading algorithms consider the first instance of buy and sell signals if there are multiple sequential signals of the same type. For instance, if there are ten back-to-back buy signals and the eleventh is a sell signal, only the first buy signal is considered and sent to the order book for execution as a limit order. The consecutive nine signals are ignored, and the holding is sold at the period when the eleventh signal is generated. Moreover, if there are any holdings at the end of a trading day, it carries over to the next day without being closed, except for the last day of trading (i.e. the end of the test dataset) when all outstanding positions are closed.

4.1. Simple Bollinger Bands Trading Strategy

The Bollinger Bands upper and lower bands are constructed two, 20-tick standard deviations around the 20-tick simple moving average of SPY Last prices. In essence, the typical Bollinger Bands are a 95% confidence interval fitted around a smoothed process. While wide bands suggest a highly volatile process, narrow bands suggest a process with muted volume. Events happening around the upper and lower bands, or thresholds, are significant, and are either used to generate signals or confirm the signals generated by other indicators. For this particular strategy, if the process crosses above the upper band then a sell signal is generated; and conversely, if the process crosses below the lower band then a buy signal is generated. The idea behind this logic model is based on the contrarian belief that a process that crosses a threshold will (mean) revert and get closer to its smoothed process or the moving average.

$$\text{Simple Moving Average} = \frac{\sum_{i=1}^{n} y_i}{n} \tag{25}$$

The Figure 5 shows the results of the Standard Bollinger Trading Strategy on October 23rd, 2020. Based on the results, the measure of the percentage difference between the upper band and the lower band is moderate. The bandwidth remains approximately constant throughout. This suggests that the volatility measured by the standard deviation is non-increasing and non-decreasing.

$$\text{Sample Standard Deviation} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})}{n-1}} \tag{26}$$

Few squeezes are observed on this graph before 16:15, at 16:33 and at 17:12. The evidence of the narrowing bands means that the volatility at those times falls to a very low level.

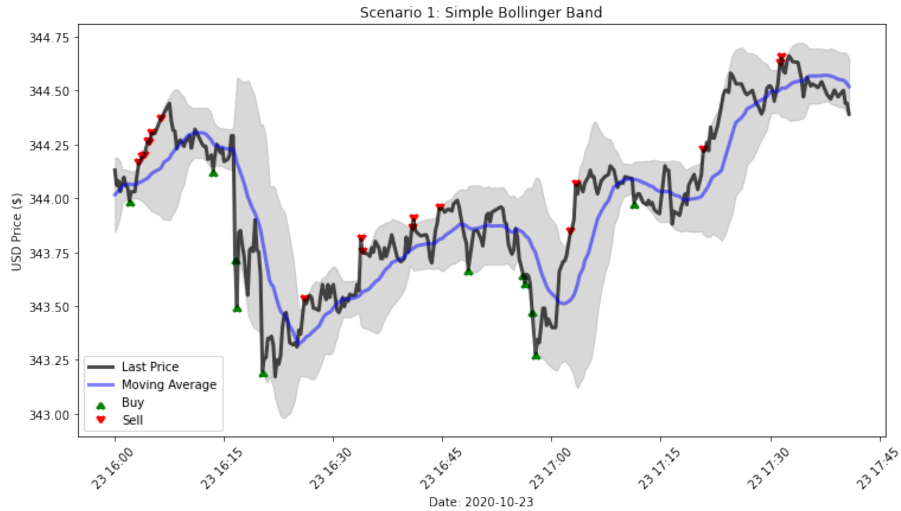


Figure 5. Simple Bollinger Bands Trading Strategy

From the pattern of the observed last price, several W-Bottoms are identified at times 16:00, 16:15, 16:20. The identification of these W-Bottoms is of interest when the second low is lower than the first but holds above the lower band as seen at time 16:20. Those signals make for a profitable trade.

4.2. Bollinger Bands + VAR

This strategy extends the Simple Bollinger Bands trading strategy discussed in 4.1. In addition to simple Bollinger Bands around the actual process, we use the predicted VAR(1) price to build a one-step forward forecast of the Bollinger Bands. The Figure 6 shows the result of this trading strategy on October 23, 2020.

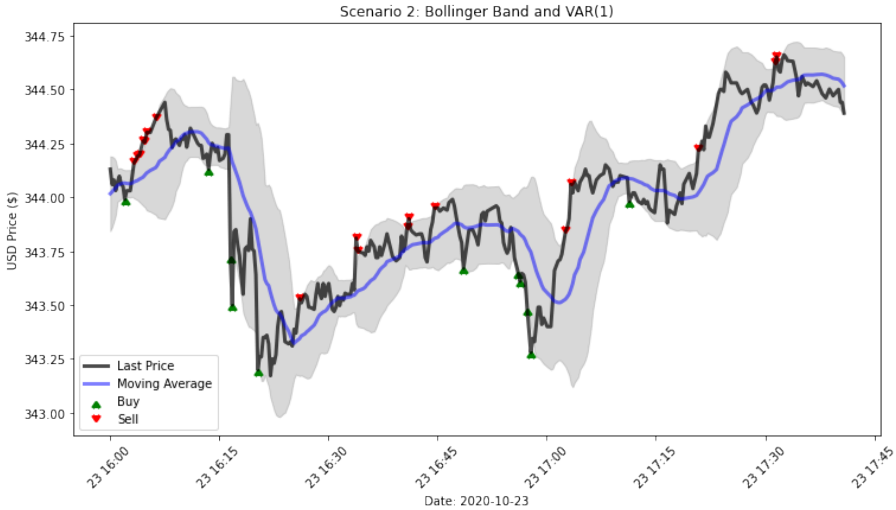


Figure 6. Bollinger Bands + VAR Trading Strategy

4.3. Bollinger Bands + VAR + 2 SD GARCH Volatility Trading Strategy

This strategy extends the Simple Bollinger Bands Trading Strategy discussed in 4.2, using predicted SPY Last prices and standard deviations from the VAR(1) and GARCH models respectively. In addition to Bollinger Bands around the actual process, we use the VAR(1) price and GARCH(1,1) volatility forecasts to build a one-step forward forecast of the Bollinger Bands. The logic model is thus extended as follows: if the actual process crosses the upper (lower) band upwards (downwards) and the one-step forward process is below (above) its respective upper (lower) band, a sell (buy) signal is generated. The Figure 7 shows the result of this trading Strategy on October 23, 2020.

4.4. Bollinger Bands + 1 Min MACD Trading Strategy

For this scenario, we use a 1-minute MACD indicator to generate trading signals and 15-second Bollinger Bands to confirm these signals. The MACD is based on an Exponential Moving Average (EMA) where older observations are given a lower weight (lower importance) for estimating the trend of a process. The EMA is given by:

$$EMA_t = ay_t + (1-a)EMA_{t-1} \tag{27}$$

The logic of the Simple Bollinger Bands is reversed to confirm the signals generated by MACD, as follows: if the MACD crosses the MACD signal line upwards (downwards) and if the process crosses the lower (upper) band upwards (downwards), then a buy (sell) signal is generated. Unlike the standard case where breaking out of upper and lower bands suggests a significant event and subsequent mean reversion, the Bollinger Bands confirm the direction of the trend.

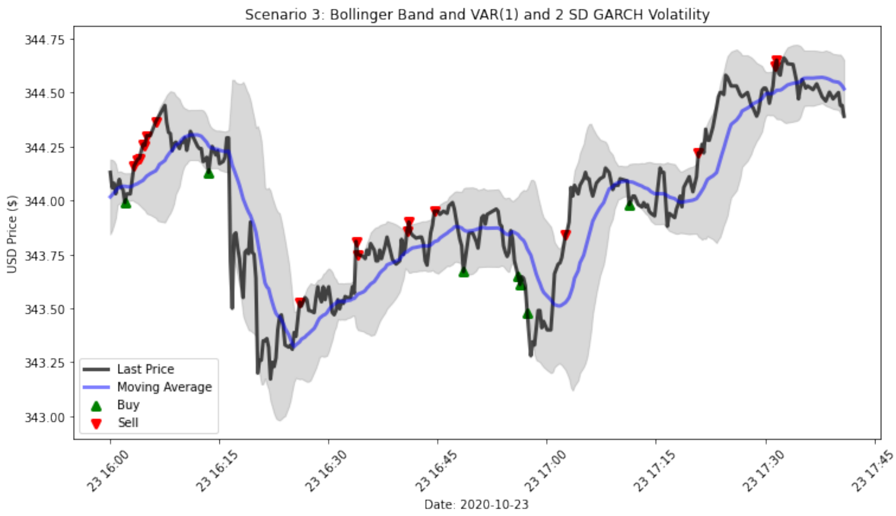


Figure 7. Bollinger Bands + VAR + 2 SD GARCH Volatility Trading Strategy

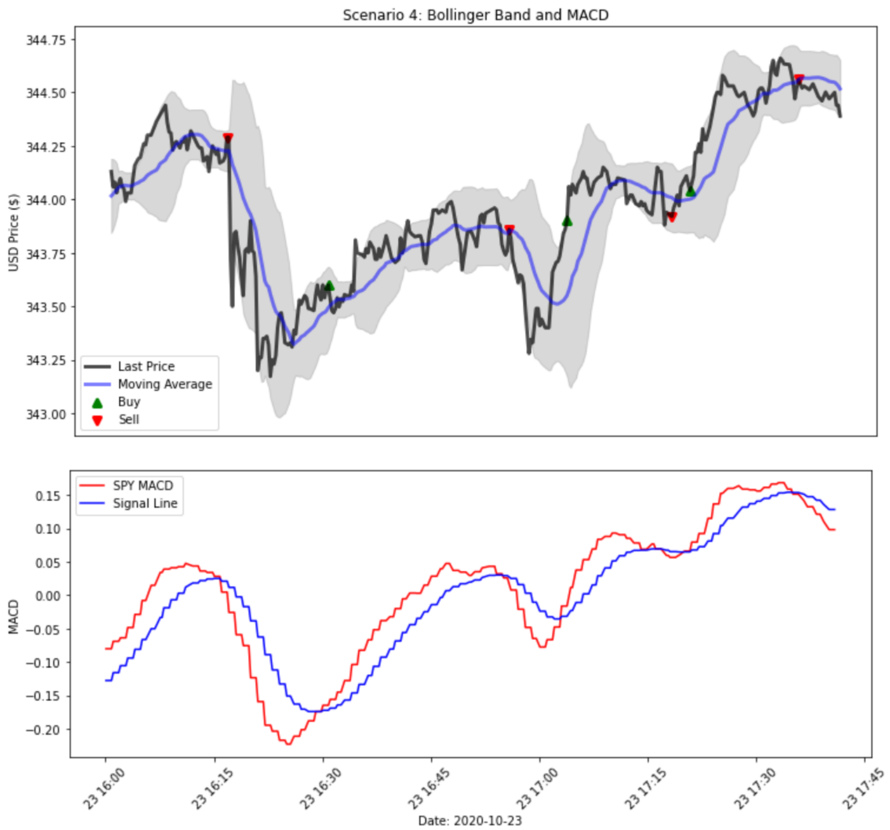


Figure 8. Bollinger Bands + VAR + 1 Min MACD Trading Strategy

4.5. Bollinger Bands + VAR + 1 Min MACD Trading Strategy

For this scenario, we use a 1-minute MACD indicator to generate trading signals and 15-second Bollinger Bands and one-step forward VAR(1) forecasts to confirm these signals. Similar to the previous scenario, the logic of the Simple Bollinger Bands is reversed to confirm the signals generated by MACD. The VAR(1) model is an additional process to verify the trend such that if the VAR(1) forecast is higher (lower) than the process, then that verifies a buy (sell) signal.

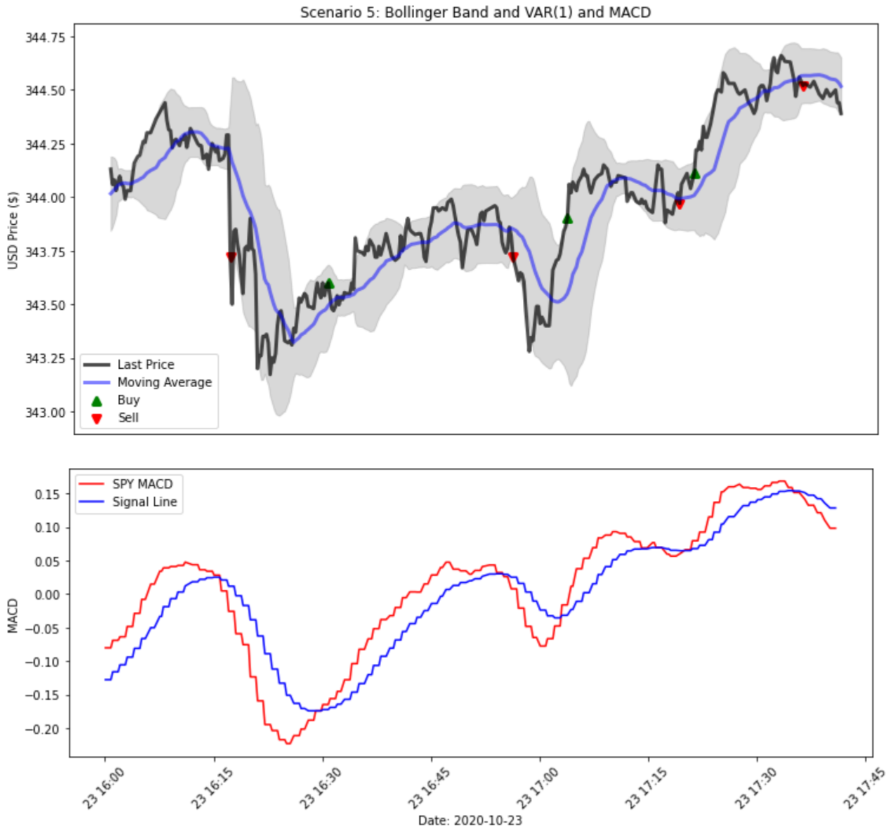


Figure 9. Bollinger Bands + 1 Min MACD Trading Strategy

5. DISCUSSION

This section examines the results obtained from the various time series forecasting models.

5.1. Discussion of the Model Results

Followed by an exploratory analysis, the results of a statistically significant Granger Causality Test indicate that the VIX Last is Granger Causing SPY Last, starting at lag 2 and continuing to lag 12. This is also inferred and supported by the time series plot of the two vectors in the Figure 2, where an increase in the VIX Last causes a decrease in the SPY Last. Next, similar to most stock price time series, the results of a KPSS test suggest that the level dataset is non-stationary with varying mean and volatility. To convert the non-stationary dataset to a stationary dataset, we differentiate the dataset. The first-difference and second-difference of the dataset are stationary. We then use the first-differenced dataset to build VAR(1), VAR(2), and VAR(3) models. We observe that on the SPY Last equation the coefficients on the first lag of SPY Last and VIX Last are statistically insignificant, while coefficients at the subsequent lags are statistically significant. In order to obtain a strong model, we second-difference the dataset and re-fit it to the VAR models, producing statistically significant coefficients. The model performance increases as the number of lags increases. In order to avoid overfitting and model misspecification and have a parsimonious model, the VAR(1) model is selected as the predictive model.

Traders most often look at a moving average of a time series to analyze how it is evolving vis-à-vis its mean and whether a trend is evolving. If a trend is evolving above or below the upper or lower Bollinger Bands, it will prompt a buy or sell decision. While not being a part of the VAR model, a moving average is integrated into an autoregressive model by way of fitting a ARIMA(2,1,1) model (or ARIMA(2,1,1) with seasonality and exogenous regressors, also known as SARIMAX(2,1,1)) on the level dataset. An exogenous variable is integrated into an ARIMA model in the form of a covariate, βX_t. However, the β is hard to interpret and has to be considered alongside a change in the conditional mean of the dependent variable. We find that the first-differenced time series of SPY Last does not have any seasonality, so we turn off the seasonality feature. Ultimately, we find that the VIX Last as an exogenous variable has no significant impact on the model. In consequence, we do not explore it further as part of ARIMA models.

Financial time series, such as the time series of a stock, generally exhibits heteroskedasticity. However, time series models like VAR and ARIMA do not capture heteroskedasticity, which means the parameters of the models will be inefficient. Fortunately, the GARCH model is a conditional heteroskedasticity model that can account for volatility clustering. The Figures 1 and 6 suggest that the dataset in question is heteroskedastic and GARCH(1,1), GARCH(2,1), and GARCH(2,2) models are fitted accordingly. Among all three models, only the GARCH(1,1) model produces statistically significant coefficients.

We produce various one-period forward prediction datasets through the fitted VAR, ARIMA, and GARCH models. We manually calculate the VAR(1), VAR(2), and VAR(3) model forecasts based on the lagged, second-differenced observations of SPY and VIX, and invert the transformations to produce level forecasts. We then compute the MSE and MAE of all three models to test the prediction accuracy against the validation dataset. As such, we find that the VAR(3) model produces the most accurate predicted values. However, for parsimony and ease of implementation, we fit a VAR(1) model for use in algorithmic trading. For the ARIMA(2,1,1) model, we use a Kalman filtering technique to train parameters on the training dataset, and then pass those parameters to a new model fitted on the test dataset. We then generate one-step and multiple-step forward forecasts. We do not have to invert transformations done for the ARIMA model, since they are done automatically by the model. Similar to the VAR models, we compute the MSE and MAE of the predicted values against the validation dataset. Finally, since the GARCH(1,1) model is fitted using the level training dataset, the one-step forward forecasts of conditional mean and variance are easily produced. We use the one-step forward conditional volatilities from GARCH(1,1) in constructing one-step forward Bollinger Bands.

5.2. Discussion of the Trading Strategies

The profitability of various algorithmic trading strategies are presented in the Table 15. An algorithmic trading strategy based on simple Bollinger Bands with a Return on Investment (ROI) of 9.42% turns out to be slightly better than a Buy-and-Hold strategy with an ROI of 8.80%. In a perfect world, where there are no additional costs to trading strategies, this trading strategy will yield good profits. In reality, however, none of the assumptions we make hold true. Therefore, if we consider the additional costs of trading in the form of fees and taxes, in addition to the costs of setting up and maintaining an algorithmic trading system, this trading strategy fares worse than a comparable investment strategy with similar ROI.

Next, we consider a trading strategy based on Bollinger Bands and VAR(1) forecasts, and notice that the VAR(1) model does not add any value versus an equivalent strategy with simple Bollinger Bands. The ROI of this strategy is exactly the same as the ROI of a strategy with simple Bollinger Bands at 9.42%. This suggests that a VAR(1) model with the SPY Last and VIX Last does not help to predict the direction of the one-step forward price on an out-of-sample dataset better than a random binary process.

For the Scenario 3, we fit Bollinger Bands twice the GARCH volatilities around the one-step forward moving averages based on a VAR(1) model. The idea behind this strategy is to confirm the movement of the one-step forward process vis-à-vis its own thresholds, rather than to confirm the actual process with a one-step forward process. This trading strategy fares slightly worse than the previous trading strategies but better than a Buy-and-Hold strategy, with an ROI of 9.37%. Overall, it does not turn out to be a lucrative strategy.

For the Scenario 4, we fit a MACD indicator on a 1-minute dataset and use the Bollinger Bands to confirm the signals generated by the MACD indicator. The idea behind introducing the MACD is that it is a smoothed process and thus a more robust measure than Bollinger Bands for representing the trend of the process. An upward (downward) trend suggests a buy (sell) signal, and the faster moving Bollinger Bands is used to confirm the trend. This strategy performs the best among all strategies, with an ROI of 13.05%. Even after costs are considered, including taxes (depending on the jurisdiction where realized gains are taxed according to the investor's marginal tax bracket), this strategy may still yield better results than a Buy-and-Hold strategy.

Finally, for the Scenario 5, we consider a combination of the VAR(1) model, Bollinger Bands and 1-minute MACD. Compared to the Scenario 4, the one-step forward VAR(1) forecasts destroys signals, as the number of signals are less than those generated for the Scenario 4, and curb overall profitability of the strategy with an ROI of 7.84%. This strategy also turns out to be the worst among all strategies.

Table 15: The Profitability of Trading Strategies

Scenario	Description	Profit ($)	ROI (%)
Benchmark	*Buy-and-Hold Investment Strategy*	*30.29*	*8.80*
1	Simple Bollinger Bands	32.50	9.42
2	Bollinger Bands + VAR	32.50	9.42
3	Bollinger Bands + VAR + 2 SD GARCH Volatility	32.30	9.37
4	Bollinger Bands + 1 Min MACD	44.89	13.05
5	Bollinger Bands + VAR + 1 Min MACD	26.98	7.84

6. CONCLUSION AND RECOMMENDATION

In this paper, we consider linear process models using the VIX Index as a proxy for market sentiment, to predict the movement of the SPY ETF prices. The coefficients on the SPY lags of the ARIMA(2,1,1) model turn out to be statistically insignificant with the addition of the VIX Index, even though this particular model performs better out-of-sample than the ARIMA(2,1,1) model without the VIX Index. Furthermore, the one-step forward forecasts generated by a VAR(1) model performs better than an ARIMA(2,1,1) model. However, in the context of algorithmic trading strategy, it counters the many signals generated by Bollinger Bands and MACD. The one-step forward GARCH(1,1) volatilities do not help in the context of algorithmic trading strategies either, perhaps because of the limited predictive capability of the VAR(1) model.

Our modeling results show that there are opportunities in other modelling structures. We believe that models that consider non-linear relationships such as clustering will perform better in forecasting stock prices. In particular, we expect that simple neural networks such as Multi Layer Perceptron (MLP) and more complicated recurrent neural networks such as Long Short-Term Memory (LSTM) will learn the jumps and diffusion in the dataset well and perform better out-of-sample. We anticipate that deep learning methods such as the Natural Language Processing (NLP) models will present enhanced capability in learning market sentiment and generating appropriate signals to be used in algorithmic trading.

Funding: "This research received no external funding"

Institutional Review Board Statement: "Not applicable"

Informed Consent Statement: "Not applicable"

Data Availability Statement:

Acknowledgments:

Conflicts of Interest: "The authors declare no conflict of interest."

Sample Availability: Samples of the compounds ... are available from the authors.

References

Amisano, Gianni and Carlo Giannini. 2012. *Topics in structural VAR econometrics*. Springer Science & Business Media.

Baker, Malcolm and Jeffrey Wurgler. 2006. Investor sentiment and the cross-section of stock returns. *The journal of Finance 61*(4), 1645–1680.

Baker, Malcolm and Jeffrey Wurgler. 2007. Investor sentiment in the stock market. *Journal of economic perspectives 21*(2), 129–152.

Bekiros, Stelios D and Dimitris A Georgoutsos. 2008. Non-linear dynamics in financial asset returns: the predictive power of the cboe volatility index. *The European Journal of Finance 14*(5), 397–408.

Bollerslev, Tim. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics 31*(3), 307–327.

Box, GEORGE EP, Gwilym M Jenkins, and G Reinsel. 1970. Time series analysis: forecasting and control holden-day san francisco. *BoxTime Series Analysis: Forecasting and Control Holden Day1970*.

Chang, Yu Sang and Dohyun Pak. 2018. Warren buffett value indicator vs. gdp size–is the relationship superlinear? *International Journal of Economics and Business Research 15*(2), 223–235.

Clarke, Roger G and Meir Statman. 1998. Bullish or bearish? *Financial Analysts Journal 54*(3), 63–72.

Corrado, Charles J and Thomas W Miller, Jr. 2005. The forecast quality of cboe implied volatility indexes. *Journal of Futures Markets: Futures, Options, and Other Derivative Products 25*(4), 339–373.

cypowered, jcoder22, yawomkobara. 2021. spyvix. https://github.com/wqucapstoneG29/spyvix.

Fernandes, Marcelo, Marcelo C Medeiros, and Marcel Scharth. 2014. Modeling and predicting the cboe market volatility index. *Journal of Banking & Finance 40*, 1–10.

Gold, Steven. 2018. Stock market algorithmic trading: A test of bollinger bands incorporating the squeeze effect and macd conditions. *Journal of Applied Financial Research 1*.

Granger, Clive WJ. 1980. Testing for causality: a personal viewpoint. *Journal of Economic Dynamics and control 2*, 329–352.

Handout, Optional TA and TA Roberto Perrelli. Introduction to arch & garch models.

Kearney, Colm and Sha Liu. 2014. Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis 33*, 171–185.

Khan, Roomy. 2021, March. Social media fueled stock market trading: The unsuspecting need to be protected. [Online; posted 8-March-2021].

Lee, Patrick. 1991. *Just how Risky are Equities Over the Long Term?*. Institute of Actuaries.

Lei, Violet UT, Simon So, and Maggie Zou. 2012. Investor sentiment-relationship between vix and trading volume. *Available at SSRN 2136802*.

Li, Feng. 2006. Do stock market investors understand the risk sentiment of corporate annual reports? *Available at SSRN 898181*.

Lütkepohl, Helmut. 2005. *New introduction to multiple time series analysis*. Springer Science & Business Media.

Mislinski, Jill. 2021. Market cap to gdp: An updated look at the buffett valuation indicator. *Advisor Perspectives*.

Nielsen, Mads R. Deviations from the covered interest rate parity.

Souza, Thársis Tuani Pinto, Olga Kolchyna, Philip C Treleaven, and Tomaso Aste. 2015. Twitter sentiment analysis applied to finance: A case study in the retail industry. *arXiv preprint arXiv:1507.00784*.