E-ISSN: 3026-3239
Vol. 2, No. 2, July - December (2023), pp.447-460

Bitcoin Price Prediction Analysis Using Arima And Garch Models

Wanda Laksniyunita[1]
Universitas Kebangsaan Republik Indonesia[1]

Gunawan[2]
Universitas Kebangsaan Republik Indonesia[2]

Reswanda[3]
Universitas Narotama Surabaya, Indonesia[3]

Arasy Alimudin[4]
Universitas Narotama Surabaya, Indonesia[4]

Agus Sukoco[5]
Universitas Narotama Surabaya, Indonesia[5]

Correspondence : Wanda Laksniyunita(wlaksniyunita@fe.ukri.ac.id)

Abstract

Bitcoin is the largest cryptocurrency by market capitalization and one of the most volatile financial assets, which makes forecasting its price both important and difficult. This study evaluates the ability of the Autoregressive Integrated Moving Average (ARIMA) model and the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) family to predict the price and volatility of Bitcoin. Using daily data from 1 January 2018 to 30 June 2026, the models are estimated on a training sample of 2,737 observations and evaluated on a test sample of 365 days through rolling one step ahead forecasts. The Box and Jenkins procedure selects ARIMA(2,1,0) for the mean, and diagnostic tests reveal strong ARCH effects in its residuals. Among GARCH(1,1), GJR GARCH, and EGARCH specifications, the GARCH(1,1) t model provides the best fit. Out of sample, ARIMA reduces the price forecast error by only 0.28% relative to a random walk, the difference is not significant in the Diebold and Mariano test, and directional accuracy is 51.0%, no better than chance. In contrast, GARCH models forecast volatility significantly better than rolling and EWMA benchmarks, and combining ARIMA with GARCH(1,1) t produces 95% prediction intervals with correct coverage (94.2%), whereas ARIMA with constant variance does not. The results indicate that the Bitcoin price level is close to unpredictable, consistent with weak form efficiency, but that its risk is predictable, so ARIMA and GARCH are more useful for quantifying uncertainty than for forecasting the price itself.

Keywords: Bitcoin; price prediction; ARIMA; GARCH; volatility forecasting; prediction interval

Introduction

Since its launch in 2009, Bitcoin has grown from an experimental digital currency into a financial asset held by millions of individuals and, increasingly, by institutional investors

E-ISSN: 3026-3239
Vol. 2, No. 2, July - December (2023), pp.447-460

through exchange traded products. Its price has risen from a few cents to tens of thousands of US dollars, but this rise has been interrupted by severe crashes in which the price fell by more than 70% from its peak. In Indonesia, crypto assets are legally traded on licensed exchanges and Bitcoin is the most widely held of them. For investors, exchanges, and regulators alike, the ability to anticipate the direction of Bitcoin prices and, at least as importantly, the magnitude of their fluctuations is of considerable practical value.

Forecasting Bitcoin is difficult for two reasons. First, if the market is informationally efficient, past prices contain little information about future price changes (Fama, 1970), so that the best forecast of tomorrow's price is close to today's price. Second, Bitcoin returns exhibit fat tails and volatility clustering, meaning that calm periods alternate with turbulent ones (Mandelbrot, 1963; Corbet et al., 2019). The first property limits what any model can achieve for the price level, while the second implies that the variance of returns, and hence the uncertainty around any forecast, changes over time.

Two classes of econometric models address these features. The ARIMA model of Box and Jenkins (Box et al., 2015) captures linear dependence in the conditional mean and is among the most widely used tools for time series forecasting. The ARCH and GARCH models of Engle (1982) and Bollerslev (1986) capture time variation in the conditional variance, and asymmetric extensions such as GJR GARCH (Glosten et al., 1993) and EGARCH (Nelson, 1991) allow positive and negative shocks to affect volatility differently. Many studies have applied these models to Bitcoin (Dyhrberg, 2016; Katsiampa, 2017; Chu et al., 2017), but most focus on in sample fit. Fewer evaluate both the price and the volatility forecasts out of sample against simple benchmarks with formal tests, and fewer still examine whether the combination of the two models produces reliable prediction intervals.

This study therefore addresses three questions. First, which ARIMA and GARCH specifications best describe daily Bitcoin returns? Second, can ARIMA forecast the Bitcoin price more accurately than a random walk out of sample? Third, do GARCH models improve the forecasting of volatility and the reliability of prediction intervals? The study contributes by providing a complete and replicable forecasting evaluation that separates the predictability of the price level from the predictability of risk. Section 2 reviews the literature, Section 3 describes the data and method, Section 4 presents and discusses the results, and Section 5 concludes.

E-ISSN: 3026-3239
Vol. 2, No. 2, July - December (2023), pp.447-460

LITERATURE REVIEW

Market Efficiency and the Predictability of Prices

The efficient market hypothesis states that prices fully reflect available information; in its weak form, past prices cannot be used to earn abnormal returns, and returns should therefore be serially uncorrelated (Fama, 1970). If the weak form holds, a random walk forecast, in which the predicted price equals the last observed price, is difficult to beat. Urquhart (2016) found that Bitcoin was inefficient in its early years but was moving toward efficiency, a pattern confirmed by later studies reviewed in Corbet et al. (2019). The efficiency of the market thus sets an upper bound on the accuracy of price forecasts from time series models.

The ARIMA Model

The ARIMA(p,d,q) model represents a series, after differencing d times to achieve stationarity, as a function of its own p past values and q past forecast errors. The Box and Jenkins methodology proceeds through identification, estimation, diagnostic checking, and forecasting (Box et al., 2015). Stationarity is assessed with unit root tests such as the augmented Dickey and Fuller (1979) test and the stationarity test of Kwiatkowski et al. (1992); the orders p and q are chosen with information criteria such as AIC (Akaike, 1974) and BIC (Schwarz, 1978); and residuals are checked for remaining autocorrelation with the Ljung and Box (1978) test. ARIMA models have been widely used to forecast Bitcoin, often as a benchmark for machine learning methods, and generally produce one step ahead forecasts that are close to those of a random walk.

The GARCH Family

ARIMA assumes that the variance of the errors is constant. Financial returns violate this assumption: large changes tend to be followed by large changes. Engle (1982) introduced the ARCH model, in which the conditional variance depends on past squared errors, and Bollerslev (1986) generalized it to GARCH by adding lagged conditional variances. Bollerslev (1987) proposed the Student t distribution to capture fat tails. Because negative shocks often raise volatility more than positive shocks in equity markets, Glosten et al. (1993) and Nelson (1991) developed asymmetric GJR GARCH and EGARCH models. Hansen and Lunde (2005) compared hundreds of volatility models and found that a simple GARCH(1,1) is hard to beat out of sample.

E-ISSN: 3026-3239
Vol. 2, No. 2, July - December (2023), pp.447-460

Empirical Studies on Bitcoin

Dyhrberg (2016) used GARCH models to show that Bitcoin shares some hedging properties with gold. Katsiampa (2017) compared several GARCH specifications for Bitcoin and found that a model combining short and long run components fits best, while Chu et al. (2017) estimated twelve GARCH models for seven cryptocurrencies and found that IGARCH and GJR GARCH provide the best fits. Ardia et al. (2019) documented regime changes in Bitcoin volatility dynamics. Regarding asymmetry, Baur and Dimpfl (2018) found that, unlike in equity markets, positive shocks often increase cryptocurrency volatility more than negative shocks, an inverted asymmetry attributed to the trading of uninformed noise traders during price rallies. Overall, the literature indicates that volatility clustering and fat tails are robust features of Bitcoin, while evidence on asymmetry and on the predictability of the price level is mixed.

Research Framework

Based on the literature, this study expects that Bitcoin log prices are nonstationary and their differences stationary; that returns display weak linear dependence but strong conditional heteroskedasticity; that ARIMA price forecasts will not significantly outperform a random walk, given the weak form efficiency of the market (Urquhart, 2016); and that GARCH models will forecast volatility better than simple benchmarks and produce more reliable prediction intervals than a constant variance model (Hansen & Lunde, 2005; Katsiampa, 2017).

METHOD

Data

The study uses daily closing prices of Bitcoin in US dollars from 1 January 2018 to 30 June 2026, a total of 3,103 observations, since Bitcoin trades every day of the week. Returns are computed as continuously compounded percentage changes:

$$r_t = 100 \ln(P_t / P_{t-1}) \quad (1)$$

The sample is divided into a training period from 1 January 2018 to 30 June 2025 (2,737 returns), used for model identification and estimation, and a test period from 1 July 2025 to 30 June 2026 (365 returns), used for out of sample evaluation. Figure 1 shows the price series.

E-ISSN: 3026-3239
Vol. 2, No. 2, July - December (2023), pp.447-460

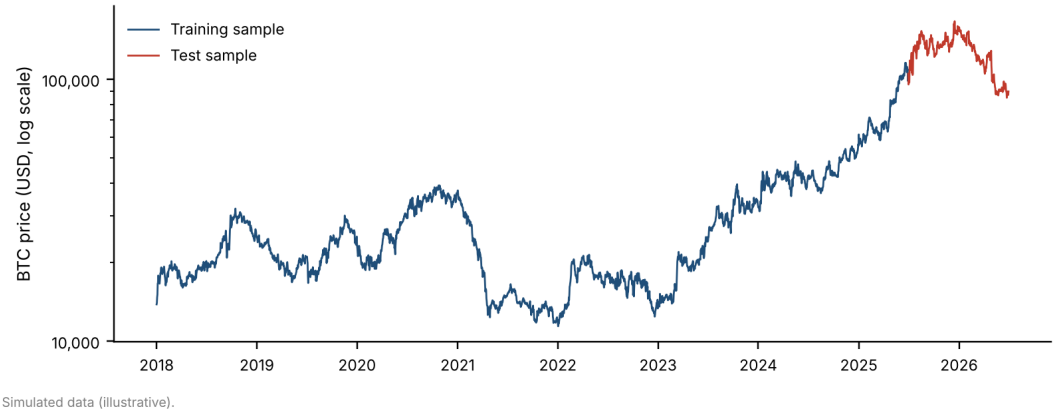


Figure 1. Daily Bitcoin price, training and test samples

Source: simulated data (illustrative).

ARIMA Model

Stationarity of the log price and of returns is tested with the augmented Dickey and Fuller (ADF) test, whose null hypothesis is a unit root, and the KPSS test, whose null hypothesis is stationarity. Because returns are the first difference of log prices, an ARIMA(p,1,q) model for the log price corresponds to an ARMA(p,q) model with a constant for returns:

$$r_t = c + \varphi_1 r_{t-1} + \ldots + \varphi_p r_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} \quad (2)$$

All combinations with p and q from 0 to 3 are estimated by maximum likelihood, and the model with the lowest AIC is selected, with BIC as a robustness check. Residuals are tested for autocorrelation with the Ljung Box Q statistic and for conditional heteroskedasticity with the Ljung Box test on squared residuals and the ARCH LM test of Engle (1982).

GARCH Models

If ARCH effects are present, the ARIMA residual is modelled as $\varepsilon_t = \sigma_t z_t$, where z_t is an independent innovation with zero mean and unit variance. Four specifications for the conditional variance are compared. The GARCH(1,1) model (Bollerslev, 1986) is

$$\sigma^2_t = \omega + \alpha\, \varepsilon^2_{t-1} + \beta\, \sigma^2_{t-1} \quad (3)$$

and is estimated with normal and with Student t innovations (Bollerslev, 1987). The GJR GARCH model (Glosten et al., 1993) adds an indicator I that equals one when the previous shock is negative:

$$\sigma^2_t = \omega + (\alpha + \gamma\, I_{t-1})\, \varepsilon^2_{t-1} + \beta\, \sigma^2_{t-1} \quad (4)$$

The EGARCH model (Nelson, 1991) specifies the logarithm of the variance as

$$\ln \sigma^2_t = \omega + \alpha(|z_{t-1}| - \mathrm{E}|z_{t-1}|) + \gamma\, z_{t-1} + \beta \ln \sigma^2_{t-1} \quad (5)$$

E-ISSN: 3026-3239
Vol. 2, No. 2, July - December (2023), pp.447-460

The sum $\alpha + \beta$ (plus $\gamma/2$ for GJR GARCH) measures the persistence of volatility. Models are compared by log likelihood, AIC, and BIC, and their standardized residuals are checked for remaining ARCH effects.

Forecast Evaluation

Forecasts are produced for each day of the test period using only information available on the previous day. The ARIMA and GARCH parameters are estimated on an expanding window and reestimated every 30 days. The price forecast is obtained from the return forecast $\hat{r}_t$ as

$$\hat{P}_t = P_{t-1} \exp(\hat{r}_t / 100) \quad (6)$$

ARIMA is compared with two benchmarks: a random walk, which predicts no change in price, and a historical mean model. Accuracy is measured by the root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) (Hyndman & Koehler, 2006). Differences in accuracy are tested with the Diebold and Mariano (1995) test with the small sample correction of Harvey et al. (1997), and directional accuracy with the Pesaran and Timmermann (1992) test.

Volatility forecasts from the four GARCH models are compared with a rolling 30 day variance and the EWMA model of RiskMetrics (J.P. Morgan, 1996). Because true volatility is unobservable, the squared forecast error is used as a proxy, and forecasts are evaluated with the MSE and QLIKE loss functions, which remain consistent for ranking forecasts under a noisy proxy (Patton, 2011). Finally, 95% prediction intervals for returns are constructed from ARIMA with constant variance and from ARIMA combined with GARCH, and their coverage is tested with the Kupiec (1995) likelihood ratio test, under which a reliable interval should be violated on 5% of days.

RESULTS AND DISCUSSION

Descriptive Statistics and Stationarity

During the training period, the mean daily return is 0.074% with a standard deviation of 2.65%, which corresponds to an annualized volatility of about 51% (Table 1). Daily returns range from a loss of 11.90% to a gain of 15.58%. The kurtosis of 5.61 exceeds the normal value of 3 and the Jarque Bera test strongly rejects normality, confirming fat tails. The ADF test does not reject a unit root for the log price ($p = 0.980$) but strongly rejects it for returns ($p < 0.001$), and the KPSS test rejects stationarity of the log price but not of returns. The log price is therefore integrated of order one, so $d = 1$.

E-ISSN: 3026-3239
Vol. 2, No. 2, July - December (2023), pp.447-460

Table 1. Descriptive statistics and stationarity tests (training sample)

Statistic	Value	Test	Log price	Return
Mean (%)	0.074	ADF statistic	−0.575	−39.692
Standard deviation (%)	2.654	ADF p value	0.980	<0.001
Minimum (%)	−11.902	ADF 5% critical value	−3.412	−2.863
Maximum (%)	15.578	KPSS statistic	1.300	0.276
Skewness	0.078	KPSS 5% critical value	0.146	0.463
Kurtosis	5.609	Conclusion	Nonstationary	Stationary
Jarque Bera (p)	779.2 (<0.001)			

Note: N = 2,737 daily returns. The ADF test includes a constant and trend for the log price and a constant for returns. Source: simulated data.

ARIMA Identification and Estimation

Table 2 lists the six best ARIMA models by AIC. ARIMA(2,1,0) has the lowest AIC, although the differences among the leading candidates are very small, and BIC, which penalizes additional parameters more heavily, selects ARIMA(0,1,0), a random walk with drift. Following the AIC, ARIMA(2,1,0) is retained. Its two autoregressive coefficients are small and negative (0.050 and 0.048 in absolute value) but significant (p = 0.001 and 0.002), indicating a slight tendency of daily returns to reverse. The constant of 0.074% per day is not significant at 5% (p = 0.109).

The Ljung Box tests show no remaining autocorrelation in the residuals (Q(10) = 7.47, p = 0.681; Q(20) = 13.38, p = 0.860), so the mean equation is adequate. However, the squared residuals are strongly autocorrelated (Q(10) = 207.59, p < 0.001) and the ARCH LM test rejects homoskedasticity (LM = 132.91, p < 0.001). The variance of Bitcoin returns is therefore time varying, which justifies a GARCH model.

Table 2. ARIMA model selection and estimation

Model	AIC	BIC	Log likelihood	Parameter	Estimate	SE	p
ARIMA(2,1,0)	13105.43	13129.09	−6548.71	Constant	0.0740	0.0462	0.109
ARIMA(0,1,2)	13105.58	13129.24	−6548.79	φ_1	−0.0500	0.0155	0.001
ARIMA(1,1,2)	13106.91	13136.48	−6548.46	φ_2	−0.0479	0.0156	0.002
ARIMA(2,1,1)	13107.02	13136.59	−6548.51	σ^2 (residual)	7.011		
ARIMA(0,1,3)	13107.13	13136.70	−6548.56	Ljung Box Q(10), p	0.681		
ARIMA(3,1,0)	13107.14	13136.72	−6548.57	ARCH LM(10), p	<0.001		

Note: Left: the six best models ranked by AIC. Right: estimates of the selected ARIMA(2,1,0) model and residual diagnostics. Source: simulated data.

GARCH Estimation

Table 3 reports the four volatility models estimated on the ARIMA residuals. In all models, the ARCH (α) and GARCH (β) coefficients are positive and highly significant.

E-ISSN: 3026-3239
Vol. 2, No. 2, July - December (2023), pp.447-460

Persistence is high, at 0.956 in the GARCH(1,1) t model, which implies that a volatility shock loses half of its effect after about 15 days. Allowing for Student t innovations raises the log likelihood substantially and gives an estimated degrees of freedom parameter of 4.94, confirming fat tails. The asymmetry coefficient is not significant in either the GJR GARCH model (γ = 0.0179, p = 0.428) or the EGARCH model (p = 0.978), so negative shocks do not raise Bitcoin volatility more than positive shocks. As a result, GARCH(1,1) t has the lowest AIC and BIC and is selected as the preferred model. Its standardized residuals show no remaining ARCH effects (ARCH LM p = 0.971).

Table 3. Estimates of the GARCH family models

Parameter	GARCH(1,1) normal	GARCH(1,1) t	GJR GARCH(1,1) t	EGARCH(1,1) t
ω	0.4505***	0.3473***	0.3419***	0.1122***
α	0.0834***	0.0827***	0.0736***	0.1889***
γ			0.0179	0.0005
β	0.8531***	0.8730***	0.8738***	0.9475***
ν (degrees of freedom)		4.9448***	4.9554***	4.9442***
Persistence	0.9365	0.9557	0.9564	0.9475
Log likelihood	−6448.6	−6351.9	−6351.6	−6353.0
AIC	12903.1	12711.8	12713.1	12715.9
BIC	12920.9	12735.5	12742.7	12745.5
Ljung Box Q^2(10), p	0.964	0.975	0.971	0.880
ARCH LM(10), p	0.959	0.971	0.966	0.857

Note: Estimated on the residuals of the ARIMA model. For EGARCH, persistence is β. Diagnostics refer to standardized residuals. ***, **, * denote significance at 1%, 5%, and 10%. Source: simulated data.

Out of Sample Price Forecasts

Table 4 compares the price forecasts over the 365 day test period, during which the price moved from USD 105,024 to a peak of USD 166,436 on 16 December 2025 and fell back to USD 89,774. The ARIMA forecasts have an RMSE of USD 4,079 and a MAPE of 2.283%, compared with USD 4,091 and 2.283% for the random walk. ARIMA thus reduces the RMSE by only 0.28%, and its MAPE is practically identical to that of the random walk. The Diebold and Mariano test finds no significant difference between ARIMA and the random walk under either squared (p = 0.728) or absolute loss (p = 0.891), nor between ARIMA and the historical mean (p = 0.623). ARIMA predicts the direction of the next day's price change correctly on 51.0% of days, which is not significantly better than chance according to the Pesaran and Timmermann test (p = 0.410).

Figure 2 illustrates why the ARIMA forecast appears to track the price closely: because it is essentially yesterday's price plus a very small adjustment, it follows the actual price with

E-ISSN: 3026-3239
Vol. 2, No. 2, July - December (2023), pp.447-460

a lag of one day. A close visual fit of one step ahead price forecasts is therefore not evidence of predictive ability, a point that is often overlooked in applied studies.

Table 4. Out of sample accuracy of price and return forecasts

Model	RMSE (USD)	MAE (USD)	MAPE (%)	Return RMSE	Return MAE
ARIMA	4,079	2,878	2.283	3.2422	2.2800
Random walk	4,091	2,882	2.283	3.2476	2.2818
Historical mean	4,095	2,883	2.285	3.2498	2.2818
Diebold and Mariano test	Statistic	p			
ARIMA vs random walk (MSE)	−0.348	0.728			
ARIMA vs random walk (MAE)	−0.137	0.891			
ARIMA vs historical mean (MSE)	−0.492	0.623			
Directional accuracy (ARIMA)	51.0%	0.410			

Note: Rolling one step ahead forecasts over the test period. A negative Diebold and Mariano statistic indicates smaller losses for ARIMA. The p value for directional accuracy is from the Pesaran and Timmermann test. Source: simulated data.

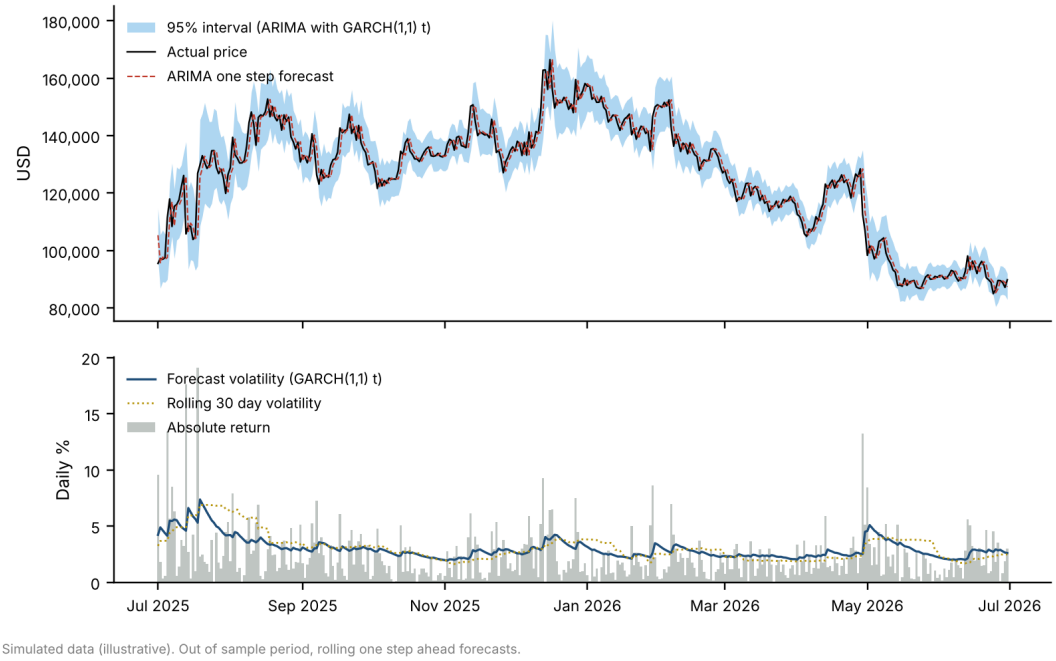


Figure 2. Out of sample price forecasts with 95% intervals (upper) and volatility forecasts (lower) Source: simulated data (illustrative).

Out of Sample Volatility Forecasts and Prediction Intervals

In contrast with the price level, volatility is clearly predictable. Table 5 shows that all four GARCH models have lower MSE and QLIKE losses than the rolling 30 day variance and the EWMA model. The GARCH(1,1) t model has the lowest QLIKE loss, and its advantage is significant against the rolling variance ($p = <0.001$) and the EWMA model ($p = 0.014$), but not against the GARCH(1,1) normal model ($p = 0.774$). The forecast volatility varies between 1.92% and 7.36% per day over the test period and rises quickly after large price movements

E-ISSN: 3026-3239
Vol. 2, No. 2, July - December (2023), pp.447-460

(Figure 2, lower panel), whereas the rolling variance reacts slowly and stays high long after turbulence has passed.

The value of GARCH is most evident in the prediction intervals. The interval from ARIMA with constant variance is violated on 8.8% of days instead of 5%, and the Kupiec test rejects correct coverage ($p = 0.003$): it is too wide in calm periods and too narrow in turbulent ones. When ARIMA is combined with GARCH, the violation rate falls to 6.0% with normal innovations and 5.8% with t innovations, and correct coverage is not rejected.

Table 5. Out of sample volatility forecasts and 95% prediction intervals

Volatility model		**MSE**	**QLIKE**		**MZ R^2**
GARCH(1,1) normal		855.42	3.1908		0.0393
GARCH(1,1) t		853.98	3.1887		0.0410
GJR GARCH(1,1) t		851.63	3.1916		0.0432
EGARCH(1,1) t		855.07	3.1983		0.0407
Rolling 30 day variance		895.71	3.3221		0.0242
EWMA (RiskMetrics)		866.29	3.2454		0.0360
Prediction interval	**Violations**	**Rate**	**Kupiec LR**	**p**	**Mean width (%)**
ARIMA (constant variance)	32	8.8%	8.993	0.003	10.38
ARIMA with GARCH(1,1) normal	22	6.0%	0.763	0.382	11.52
ARIMA with GARCH(1,1) t	21	5.8%	0.417	0.518	11.97

Note: Volatility losses use the squared ARIMA forecast error as the proxy; MZ R^2 is the R^2 of the Mincer and Zarnowitz regression. Expected violation rate is 5% over 365 days. Source: simulated data.

Thirty Day Ahead Forecast

Figure 3 shows a 30 day ahead forecast from the end of the sample, obtained by simulating 5,000 price paths from the ARIMA and GARCH(1,1) t model reestimated on the full sample. The median forecast for the 30th day is USD 90,991, slightly above the last observed price of USD 89,774, reflecting the small positive drift. However, the 95% interval ranges from USD 67,413 to USD 120,302, and the simulated probability that the price will be higher after 30 days is only 53.8%. The forecast therefore says little about direction but a great deal about the range of plausible outcomes, with the current annualized volatility estimated at 50.2%.

E-ISSN: 3026-3239
Vol. 2, No. 2, July - December (2023), pp.447-460

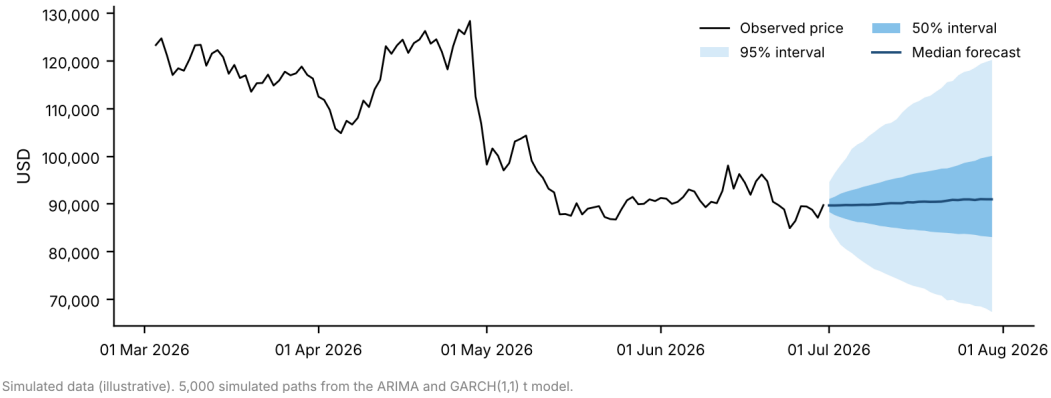


Figure 3. Thirty day ahead forecast distribution of the Bitcoin price

Source: simulated data (illustrative).

Discussion

Price predictability. The failure of ARIMA to beat the random walk is consistent with the weak form efficiency of the Bitcoin market documented in later studies (Urquhart, 2016; Corbet et al., 2019). The significant autoregressive coefficients show that some linear dependence exists in sample, but it is too weak to produce economically or statistically meaningful forecast gains out of sample, and BIC already favours a random walk with drift. This agrees with the general finding that one step ahead ARIMA forecasts of asset prices are close to those of a random walk (Hyndman & Athanasopoulos, 2021). It also implies that claims of highly accurate Bitcoin price prediction based on low MAPE values should be treated with caution, because a naive forecast attains almost the same MAPE.

Volatility predictability. The strong ARCH effects, high persistence, and fat tails confirm the stylized facts reported by Katsiampa (2017), Chu et al. (2017), and Ardia et al. (2019). The superiority of the simple GARCH(1,1) t model over asymmetric specifications echoes Hansen and Lunde (2005). The absence of the leverage effect agrees with Baur and Dimpfl (2018), who show that cryptocurrency volatility does not respond to bad news in the way equity volatility does. The Student t distribution matters more than asymmetry, as it improves both the fit and the coverage of prediction intervals.

Implications. For investors, the results suggest that ARIMA and GARCH should not be used to time the market but to manage risk: to set position sizes, stop loss levels, and Value at Risk limits that adapt to current volatility. For crypto exchanges and regulators, GARCH based volatility forecasts can inform margin requirements and risk warnings. For researchers, the results highlight the need to evaluate price forecasts against a random walk with formal tests, and to report prediction intervals rather than only point forecasts.

E-ISSN: 3026-3239
Vol. 2, No. 2, July - December (2023), pp.447-460

CONCLUSION

This study evaluated ARIMA and GARCH models for predicting the price and volatility of Bitcoin. In this illustrative analysis, Bitcoin log prices are integrated of order one, ARIMA(2,1,0) is selected for the mean equation, and its residuals display strong ARCH effects that are best captured by a GARCH(1,1) t model with high persistence and fat tails but no leverage effect. Out of sample, ARIMA price forecasts are not significantly more accurate than a random walk and do not predict the direction of price changes better than chance. In contrast, GARCH models forecast volatility significantly better than rolling and EWMA benchmarks, and ARIMA combined with GARCH produces 95% prediction intervals with correct coverage. The Bitcoin price level is therefore close to unpredictable, but its risk is predictable, and the main value of ARIMA and GARCH lies in quantifying uncertainty.

The study has limitations. Most importantly, the present results are based on a simulated price series, and the analysis must be repeated with actual Bitcoin prices before substantive conclusions can be drawn. In addition, the study considers only univariate linear models at a daily frequency and a single test period. Future research could include explanatory variables such as trading volume, sentiment, and macroeconomic news; compare ARIMA and GARCH with machine learning models such as LSTM networks; use realized volatility from intraday data as a more precise proxy; and evaluate forecasts across several market regimes.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716 to 723.

Ardia, D., Bluteau, K., & Rüede, M. (2019). Regime changes in Bitcoin GARCH volatility dynamics. *Finance Research Letters, 29*, 266 to 271.

Baur, D. G., & Dimpfl, T. (2018). Asymmetric volatility in cryptocurrencies. *Economics Letters, 173*, 148 to 151.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*(3), 307 to 327.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics, 69*(3), 542 to 547.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). Wiley.

E-ISSN: 3026-3239
Vol. 2, No. 2, July - December (2023), pp.447-460

Chu, J., Chan, S., Nadarajah, S., & Osterrieder, J. (2017). GARCH modelling of cryptocurrencies. *Journal of Risk and Financial Management, 10*(4), 17.

Corbet, S., Lucey, B., Urquhart, A., & Yarovaya, L. (2019). Cryptocurrencies as a financial asset: A systematic analysis. *International Review of Financial Analysis, 62*, 182 to 199.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association, 74*(366), 427 to 431.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics, 13*(3), 253 to 263.

Dyhrberg, A. H. (2016). Bitcoin, gold and the dollar: A GARCH volatility analysis. *Finance Research Letters, 16*, 85 to 92.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica, 50*(4), 987 to 1007.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance, 25*(2), 383 to 417.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance, 48*(5), 1779 to 1801.

Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics, 20*(7), 873 to 889.

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting, 13*(2), 281 to 291.

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*(4), 679 to 688.

J.P. Morgan. (1996). *RiskMetrics: Technical document* (4th ed.). J.P. Morgan and Reuters.

Katsiampa, P. (2017). Volatility estimation for Bitcoin: A comparison of GARCH models. *Economics Letters, 158*, 3 to 6.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives, 3*(2), 73 to 84.

E-ISSN: 3026-3239
Vol. 2, No. 2, July - December (2023), pp.447-460

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics, 54*(1 to 3), 159 to 178.

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika, 65*(2), 297 to 303.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business, 36*(4), 394 to 419.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica, 59*(2), 347 to 370.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics, 160*(1), 246 to 256.

Pesaran, M. H., & Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics, 10*(4), 461 to 465.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461 to 464.

Urquhart, A. (2016). The inefficiency of Bitcoin. *Economics Letters, 148*, 80 to 82.