Estimating and Forecasting Cryptocurrency Prices: An ARIMA-GARCH Approach

A Case Study of Dogecoin

Shaghayegh Dianat

Sharif University

February 1, 2025

Introduction

- **Background:** Dogecoin was created in December 2013, it has a 1-minute block time, making transactions faster than Bitcoin (10-minute block time).Dogecoin transactions typically have very low fees, making it suitable for microtransactions and tipping.
- **Problem:** Cryptocurrency markets are often volatile, making price prediction challenging.
- **Objective:** To analyze and forecast Dogecoin prices using time series methodology, specifically ARIMA-GARCH models.
- **Data:** Daily closing prices of Dogecoin from December 1st, 2021 to November 1st, 2023, for model estimation, with data from November 17th, 2023 to December 2nd, 2023 for forecasting.
- **Data Source:** Yahoo Finance

Literature Review

- **Previous Research:**
  - Various techniques have been used for Cryptocurrency price prediction, including machine learning and regression.
  - Statistical methods like ARIMA, AR, MA, and GARCH are frequently applied .
  - Studies have employed ARIMA models with varying degrees of accuracy.
  - GARCH models have also been used to forecast Dogecoin prices.
  - Machine learning techniques have been used, with statistical methods sometimes outperforming them.
  - Deep learning methods have shown high accuracy in Cryptocurrency price prediction.
  - Hybrid models have been used as well, combining deep learning with GARCH .
- **This Study's Focus:** Using a hybrid ARIMA-GARCH model for improved forecasting .

Methodology - ARIMA

- **ARIMA Model:** The Box-Jenkins ARIMA model (p, d, q) is used for time series analysis.
  - p: number of autoregressive (AR) terms
  - d: number of differences taken
  - q: number of moving average (MA) terms
- **Equation:**
  $(1-B)^d y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + W_t + \theta_1 W_{t-1} + \ldots + \theta_q W_{t-q}$
- **Key Assumption:** Constant variance of data:
  $W_t \sim WN(0, \sigma^2)$

Methodology - GARCH

- **GARCH Model:** Used for modeling time-varying volatility.
- **GARCH (q, p) Model:**
  - $y_t = \mu_t + r_t$
  - $r_t = \sigma_t \varepsilon_t$ where $\varepsilon_t \sim N(0,1)$
  - $\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} (\alpha_i * r_{t-i}^2) + \sum_{j=1}^{p} (\beta_j * \sigma_{t-j}^2)$
  - σ_t^2 is the conditional variance of y_t
- **Purpose:** To capture the volatility clustering often seen in financial time series .
- **Constraints:**

$$\alpha_0 > 0$$

$$\alpha_i \geq 0, \quad \text{for } i = 1, 2, \dots, q$$

$$\beta_j \geq 0, \quad \text{for } j = 1, 2, \dots, p$$

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$$

Methodology - Hybrid ARIMA-GARCH

- **Two-Stage Approach:**
  1. **ARIMA Stage:** Apply the best ARIMA model to fit stationary and linear data .
  2. **GARCH Stage:** Use the GARCH model to capture non-linear residual patterns from the ARIMA stage.
- **Benefits:** Combines linear and non-linear modeling for better forecasting .

Descriptive statistics of daily Doge price

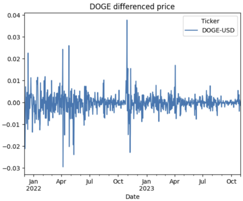
Statistic	Price
Mean	0.090638
Median	0.076609
Maximum	0.209726
Minimum	0.053012
Std. dev.	0.034476
Skewness	1.305805
Kurtosis	0.598668
Jarque-Bera	208.089654
Probability	0.0
Sum Ticker	63.446734
Sum Sq. dev	0.830822
Observations	700

Table: Source: From data and calculations on Google Colab by the author.

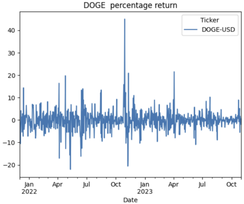
Descriptive statistics of percentage return of Doge

Statistic	RPrice
Mean	-0.053578
Median	-0.053443
Maximum	44.943197
Minimum	-22.028628
Std. dev.	4.681343
Skewness	1.333395
Kurtosis	15.300767
Jarque-Bera	6920.209518
Probability	0.0
Sum Ticker	37.451026
Sum Sq. dev	15296.651067
Observations	699

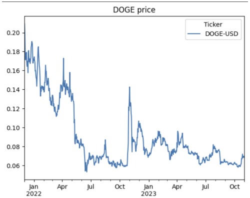
Table: Source: From data and calculations on Google Colab by the author.



(a) DOGE Differenced Price



(b) DOGE Percentage Return



(c) DOGE Price

Figure: Data Visualization

ADF Test

The Augmented Dickey-Fuller (ADF) test is a statistical test used to determine whether a time series is stationary. Stationarity is a crucial assumption in many time series models, such as ARIMA.

The ADF test estimates the following regression model:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + \epsilon_t$$

Where:

- Δy_t = the difference of the series at time t (i.e., $y_t - y_{t-1}$)
- α = a constant (drift term)
- βt = a time trend (optional)
- γy_{t-1} = the lagged level of the series
- $\phi_i \Delta y_{t-i}$ = lagged differences of the series (to account for autocorrelation)
- ϵ_t = the error term
- **Null Hypothesis (H_0)**: The time series has a unit root (i.e., it is non-stationary).

ADF Test

The ADF test statistic is calculated as:

$$ADF = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$$

Where:

- $\hat{\gamma}$ = the estimated coefficient for y_{t-1}
- $SE(\hat{\gamma})$ = the standard error of $\hat{\gamma}$

If the ADF test statistic is more negative than the critical value, the null hypothesis is rejected, indicating that the series is stationary. If the ADF test statistic is less negative than the critical value, the null hypothesis cannot be rejected, suggesting that the series is non-stationary.

ADF Test Results

```
ADF Statistic of prices: -2.5172247363641223
p-value: 0.111
1% Critical Value: -3.439960610754265
5% Critical Value: -2.8657809735786244
10% Critical Value: -2.569028437390806
Series is non-stationary so apply differencing ( or use returns).
```

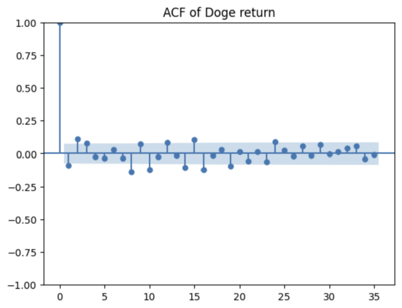
Figure: ADF Test Results for Prices

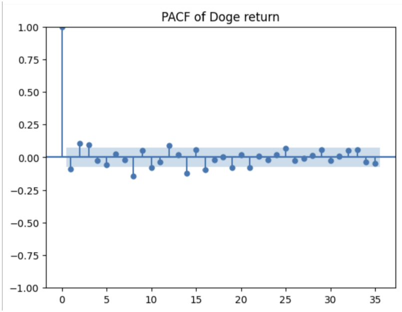
```
ADF Statistic of returns: -7.707208313285913
p-value: 0.0
1% Critical Value: -3.439960610754265
5% Critical Value: -2.8657809735786244
10% Critical Value: -2.569028437390806
```

Figure: ADF Test Results for Returns

Lags(k)	AC	PAC	Q-stat	Prob
1	-0.0881	-0.0881	5.4476	0.0196
2	0.1101	0.1032	13.9731	0.0009
3	0.0757	0.0953	18.0123	0.0004
4	-0.0251	-0.0228	18.4580	0.0010
5	-0.0339	-0.0586	19.2707	0.0017
6	0.0305	0.0215	19.9290	0.0029
7	-0.0375	-0.0184	20.9226	0.0039
8	-0.1356	-0.1441	33.9522	0.0000
9	0.0735	0.0517	37.7896	0.0000
10	-0.1210	-0.0761	48.1983	0.0000
11	-0.0237	-0.0332	48.5970	0.0000
12	0.0862	0.0890	53.9006	0.0000
13	-0.0143	0.0178	54.0472	0.0000
14	-0.1051	-0.1237	61.9527	0.0000
15	0.1031	0.0567	69.5622	0.0000
16	-0.1211	-0.0943	80.0831	0.0000
17	-0.0122	-0.0182	80.1903	0.0000
18	0.0293	0.0022	80.8090	0.0000
19	-0.0971	-0.0761	87.6059	0.0000
20	0.0118	0.0197	87.7060	0.0000
21	-0.0582	-0.0759	90.1551	0.0000
22	0.0121	0.0089	90.2609	0.0000
23	-0.0599	-0.0191	92.8585	0.0000
24	0.0881	0.0179	98.4959	0.0000
25	0.0237	0.0649	98.9050	0.0000
26	-0.0187	-0.0263	99.1600	0.0000
27	0.0550	-0.0091	101.3620	0.0000
28	-0.0109	0.0114	101.4495	0.0000
29	0.0671	0.0584	104.7439	0.0000
30	-0.0045	-0.0270	104.7585	0.0000
31	0.0150	0.0055	104.9239	0.0000
32	0.0383	0.0523	106.0026	0.0000
33	0.0550	0.0563	108.2307	0.0000
34	-0.0420	-0.0333	109.5328	0.0000
35	-0.0106	-0.0435	109.6155	0.0000

ACF and PACF of Doge Return





p	q	AIC
16	15	4099.573
10	16	4100.132
8	14	4101.548
8	16	4102.133
10	14	4102.657
16	16	4102.931
16	14	4103.101
15	3	4103.466
14	3	4103.792
10	15	4105.856
14	16	4106.130
8	15	4106.320
16	3	4106.488
3	10	4106.686
15	16	4107.880
14	14	4109.942
3	14	4110.433
8	3	4111.204
3	16	4111.372
16	10	4111.495
15	8	4111.608
14	8	4111.767
15	14	4111.932
3	15	4112.236
3	8	4112.739
16	8	4113.035
14	10	4113.822
8	8	4114.212
10	10	4114.228
10	3	4115.750
15	15	4116.490
8	10	4116.536
14	15	4116.937
10	8	4117.091
15	10	4123.264
3	3	4125.125

Results - ARIMA Model

- **Optimal ARIMA Model:** ARIMA(16,1,15) based on the AIC.
- **Check heteroscedasticity:** Check heteroscedasticity: Before building the GARCH model for the Dogecoin return series, the first step is to check whether the variance of the residuals of the ARIMA(16,1,15) model is explained by ARCH effects.
- **ARCH-LM Test:** The null hypothesis (H_0) of the ARCH LM test is that there are no ARCH effects in the residuals

Results - GARCH Model

```
ARCH Test Results
             Statistic   Value
ARCH-LM Test Statistic 44.7657
            LM p-value  0.0089
           F-statistic  1.8440
             F p-value  0.0077

Reject the null hypothesis: Significant ARCH effects detected.
```

Figure: ARCH-LM Test Results for Residual

ARCH-LM Test Indicated the presence of ARCH effects in the residuals of the ARIMA model.

- **GARCH(1,1) Model:** Used to model the volatility of Dogecoin returns .
- **Model Estimation:** Coefficients for the ARIMA(16,1,15)-GARCH(1,1) were estimated.

Variable	Coefficient	Std. Error	z-statistic	Prob.
const	-0.0527	0.1526	-0.3453	0.7299
ar.L1	0.1043	0.1456	0.7164	0.4737
ar.L2	0.2515	0.1172	2.1452	0.0319
ar.L3	0.2779	0.0878	3.1667	0.0015
ar.L4	0.1687	0.1048	1.6095	0.1075
ar.L5	0.1050	0.0988	1.0632	0.2877
ar.L6	-0.1275	0.0981	-1.2993	0.1938
ar.L7	-0.0217	0.0985	-0.2202	0.8257
ar.L8	-0.4147	0.0855	-4.8532	0.0000
ar.L9	-0.1092	0.0977	-1.1182	0.2635
ar.L10	-0.2779	0.0906	-3.0687	0.0021
ar.L11	-0.0476	0.1047	-0.4547	0.6493
ar.L12	0.1480	0.0937	1.5791	0.1143
ar.L13	0.3915	0.0842	4.6466	0.0000
ar.L14	0.2626	0.1086	2.4174	0.0156
ar.L15	-0.3613	0.1139	-3.1736	0.0015
ar.L16	-0.2648	0.0422	-6.2765	0.0000
ma.L1	-0.1728	0.1452	-1.1906	0.2338
ma.L2	-0.1290	0.1231	-1.0481	0.2946
ma.L3	-0.2202	0.0856	-2.5720	0.0101
ma.L4	-0.1851	0.1066	-1.7367	0.0824
ma.L5	-0.1682	0.1019	-1.6504	0.0989
ma.L6	0.1122	0.1016	1.1043	0.2695
ma.L7	-0.0779	0.1047	-0.7440	0.4569
ma.L8	0.3341	0.0886	3.7723	0.0002
ma.L9	0.1330	0.1032	1.2884	0.1976
ma.L10	0.2673	0.0960	2.7854	0.0053
ma.L11	0.1137	0.1086	1.0470	0.2951
ma.L12	0.0053	0.1010	0.0525	0.9582
ma.L13	-0.4213	0.0904	-4.6627	0.0000
ma.L14	-0.3661	0.1236	-2.9628	0.0030
ma.L15	0.4812	0.1324	3.6347	0.0003
sigma2	18.8095	0.8385	22.4333	0.0000
omega	1.4275	1.1926	1.1970	0.2313
alpha[1]	0.1406	0.0728	1.9328	0.0533
beta[1]	0.7893	0.1198	6.5890	0.0000

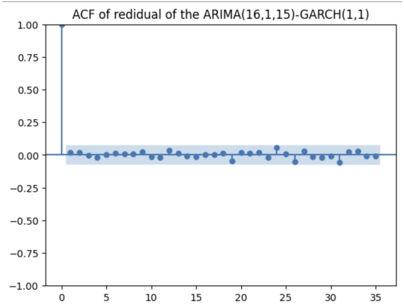
Model Evaluation

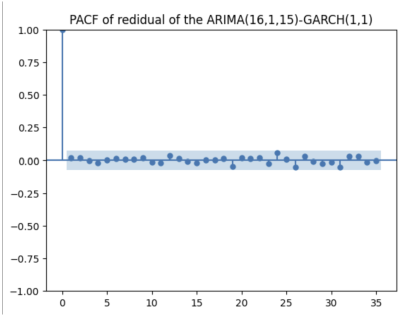
- **Residual Analysis:** Autocorrelation and partial autocorrelation of residuals were used to validate the hybrid model.
- **Ljung-Box Test:** The Ljung-Box test, is a statistical test used to determine whether there are significant autocorrelations in a time series. it is commonly used to check the adequacy of time series models, such as ARIMA models. If the model is adequate, the residuals should resemble white noise, meaning there should be no significant autocorrelations.
- **ARCH-LM test:** The results showed that the ARIMA(16,1,15)-GARCH(1,1) model has modeled the heteroscedasticity of the Dogecoin return prices series.

ACF/PACF of Residuals

Lag	AC	PAC	Q-stat	Prob
1	0.016233	0.016257	0.184993	0.667117
2	0.017371	0.017162	0.397140	0.819902
3	-0.002116	-0.002684	0.400291	0.940182
4	-0.017560	-0.017893	0.617681	0.961079
5	0.001109	0.001770	0.618549	0.987138
6	0.015349	0.016066	0.785123	0.992467
7	0.007911	0.007364	0.829440	0.997132
8	0.007544	0.006523	0.869793	0.998945
9	0.021745	0.021675	1.205596	0.998799
10	-0.015064	-0.015655	1.366988	0.999293
11	-0.016708	-0.017029	1.565809	0.999531
12	0.033061	0.034887	2.345411	0.998663
13	0.015177	0.015487	2.509934	0.999203
14	-0.007947	-0.010867	2.555114	0.999636
15	-0.016221	-0.018099	2.743606	0.999770
16	0.001928	0.004364	2.746273	0.999907
17	0.003880	0.005409	2.757087	0.999962
18	0.013486	0.012084	2.887957	0.999979
19	-0.044748	-0.047404	4.330877	0.999805
20	0.017514	0.019426	4.552246	0.999868
21	0.015697	0.016088	4.730320	0.999917
22	0.021256	0.021467	5.057328	0.999932
23	-0.021581	-0.023428	5.394923	0.999944
24	0.055538	0.058115	7.634024	0.999385
25	0.007116	0.006076	7.670838	0.999653

ACF/ PACF of Residual





Diagnostic Checking

- **ARCH-LM Test** shows that there is no ARCH effect in the squared residuals.

```
ARCH Test Results
             Statistic  Value
ARCH-LM Test Statistic 1.3555
            LM p-value 1.0000
           F-statistic 0.0432
             F p-value 1.0000

Fail to reject the null hypothesis: No significant ARCH effects.
```

- **Ljung-Box Test** shows that there is no significant autocorrelation in the residuals.

```
Ljung-Box p-value: 0.9994
```

- **AIC/BIC of Models**

```
AIC ARIMA(16,1,15): 4099.57310429001
BIC ARIMA(16,1,15): 4249.711578783726
AIC ARIMA(16,1,15)-GARCH(1,1): 3832.8825340058283
BIC ARIMA(16,1,15)-GARCH(1,1) : 3860.180438459231
```

DOGE Price Forecast Using ARIMA-GARCH

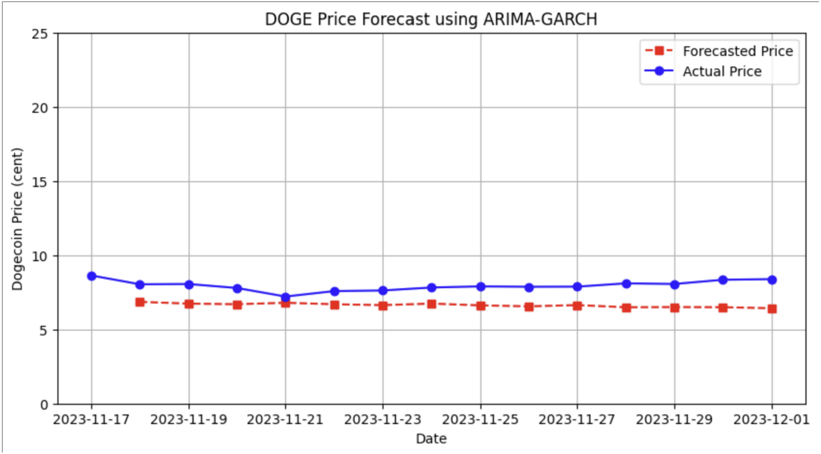


Figure: The graph shows the forecasted price (red dashed line with squares) and the actual price (blue solid line with circles) of Dogecoin in cent from November 18, 2023, to December 1, 2023.

Model Comparison

- **Model Fit:** The ARIMA(16,1,15)-GARCH(1,1) model had the smallest AIC and BIC values, indicating a better fit than the single ARIMA model.
- **Forecasting Ability:**

Model	RMSE	MAE	MAPE
ARIMA(16,1,15)-GARCH(1,1)	0.01327	0.01271	15.96060
ARIMA(16,1,15)	0.01389	0.01334	16.75911

Table: Performance Metrics of Different Models

- **Conclusion:** The ARIMA-GARCH combined model outperforms the single ARIMA model in modeling and forecasting Dogecoin prices.

Conclusion

- **Hybrid Model:** The ARIMA-GARCH hybrid model is effective for analyzing and forecasting Dogeoin prices due to high volatility.
- **Short-Term Forecasting:** The model is most suitable for short-term forecasts due to the assumption that future conditions will be similar to the past.
- **Model Flexibility:** Users need to be flexible in model selection to avoid missing other meaningful models.

Limitations and Future Work

- **Macro Factors:** The model does not account for macroeconomic factors such as GDP growth and interest rates.
- **Model Parameters:** The model is sensitive to the selection of lags and orders for both ARIMA and GARCH. Thorough testing is needed to confirm assumptions of the model, such as using AIC, BIC, and cross-validation .
- **Volatility:** The highly volatile nature of Dogecoin requires frequent recalibration of the model using techniques like rolling windows or online learning .
- **Further Research:** Future studies should compare the ARIMA-GARCH model with other forecasting models.

Applying the Model to Other Cryptocurrencies

- **Data:** Use daily closing prices of the cryptocurrency .
- **Stationarity:** Test for stationarity and apply differencing (or use return) if needed.
- **Model Selection:** Based on the ACF and PACF plots, you can identify candidate values for p and q, then use AIC to select optimal ARIMA lags.
- **Volatility:** Test for ARCH effect by use of ARCH-Lm Test,Use GARCH(1,1) model for volatility.
- **Validation:** Use Jarque-Bera, Ljung-Box, and ARCH-LM tests to validate your model.
- **Hybrid Approach:** Combine ARIMA and GARCH for linear and non-linear effects.
- **Calibration:** Due to volatility, continuously recalibrate the model.
- **Limitations:** Recognize that the model is limited to short-term forecasts.

References

Quang Phung Duy et al. (2024). Estimating and forecasting bitcoin daily prices using ARIMA-GARCH models. *Business Analyst Journal*, 45(1), pp. 11-23.

Ashley, R. A., & Patterson, D. M. (2010). A test of the GARCH (1, 1) specification for daily.

Bakar, N. A., & Rosbi, S. (2018). Weighted moving average of forecasting method for predicting bitcoin share price using high frequency data: A statistical method in financial cryptocurrency technology. *International Journal of Advanced Engineering Research and Science*, 5(1), 64–69. doi:10.22161/ijaers.5.1.11.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327. doi:10.1016/0304-4076(86)90063-1.

Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique*, 55(2), 163–172. doi:10.2307/1403192.

Katsiampa, P. (2017). Volatility estimation for bitcoin: A comparison of GARCH models. *Economics Letters*, 158, 3–6. doi:10.1016/j.econlet.2017.06.023.

MacKinnon, J. G. (1996). Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics*, 11(6), available at: https://www.jstor.org/stable/2285154.

Roy, S., Nanjiba, S., & Chakrabarty, A. (2018). Bitcoin price forecasting using time series analysis. *2018 21st International Conference on Computer and Information Technology (ICCIT)*, 1-6. doi:10.1109/ICCIT.2018.236642.

Sadorsky, P. (2006). Modeling and forecasting petroleum futures volatility. *Energy Economics*, 28(4), 467–488. doi:10.1016/j.eneco.2006.04.005.

Sarkodie, S. A., Amani, M. A., Ahmed, M. Y., & Owusu, P. A. (2023). Assessment of Bitcoin carbon footprint. *Sustainable Horizons*, 7, 100060. doi:10.1016/j.horiz.2023.100060.

Sarkodie, S. A., Ahmed, M. Y., & Owusu, P. A. (2022). COVID-19 pandemic improves market signals of cryptocurrencies–evidence from Bitcoin, Bitcoin Cash, Ethereum, and Litecoin. *Finance Research Letters*, 44(c). doi:10.1016/j.frl.2021.102049.