Forecasting the Price Volatility of Bitcoin Using a Hybrid ARIMA-GARCH Model – Project Proposal

Problem Statement and Background:

Bitcoin (BTC) has established itself as a significant and unique asset in today's world, being seen as the first of its kind in a world where thousands of cryptocurrencies now exist. This digital asset was first introduced to the world in 2009, in which an anonymous entity under the pseudonym of "Satoshi Nakamoto" published the famous Bitcoin whitepaper, creating an asset that's "fully peer-to-peer, with no trusted third party" (Investopedia 2024). This showcases the decentralised nature of Bitcoin, one of its many unique characteristics. Other notable characteristics include its limited supply (maximum supply of 21 million), its use of blockchain technology (transactions get locked in an unbreakable chain of math) and the ability for transactions to be made without being linked to personal data make it stand out against other similar assets (e.g., gold) (GX BLOCKS 2022).

However, despite emerging as a revolutionary and globally adopted asset, Bitcoin's extreme price volatility continues to present challenges for investors. Traditional models such as Geometric Brownian Motion (GBM) struggle to account for Bitcoin's non-linear and highly volatile nature characterised by volatility clustering (high/low volatility periods are prolonged and clustered together), extreme price swings (sudden rallies and declines) and persistent momentum effects. This shows how techniques used on assets such as gold or stocks cannot be used for Bitcoin as they are generally inadequate to account for Bitcoin's erratic behaviour. This presents gaps within traditional models, leaving investors with a limited set of tools to accurately predict prices and manage risk. This causes investors to rely on speculation rather than data-driven strategy, intensifying risks in an already unpredictable market (Reiff 2024).

To address these shortcomings, this project proposes a hybrid ARIMA-GARCH model designed to:

- Apply Auto-Regressive Integrated Moving Average (ARIMA) to identify linear trends
- Employ Generalised Autoregressive Conditional Heteroskedasticity (GARCH) to model Bitcoins tendency to go through periods of high and low volatility

This hybrid approach works better for Bitcoin than techniques such as GBM as it combines the best of both worlds. ARIMA handles the clear price trends while GARCH handles periods where volatility seems to cluster together, matching Bitcoin's behaviour much closer. When we bring these two methods together, we can predict Bitcoin's short-term price swings and simulate where prices might head next much better as it addresses key blind spots that other models miss. This allows traders to estimate Bitcoin's future price movements, providing improvements to managing risk and less dependence on speculation in famously unstable market.

Methodology:

The proposed project implements a hybrid ARIMA-GARCH model, aiming to accurately predict Bitcoin's price volatility and future price trends. The calculations and data analysis will be done within Python. To do so, several crucial steps are needed to be completed. The proposed methodology is as follows:

1. **Gathering historical market data**
   Historic Bitcoin data will be gathered from Yahoo Finance or other reputable sources such as CoinMarketCap. This data will be the daily closing price of Bitcoin in USD, gathered from the January 1st, 2023, to current times to mitigate outliers present and to get sufficient coverage within our data.

2. **Preparing data for analysis**
   This step will ensure that all values within the data are suitable to be used within our models. This can be done with the following steps taken:
   - Handle missing values and irregularities within the dataset (interpolation/removal).
   - Convert prices to log returns for stationarity (necessary transformation to stabilise variance). An Augmented Dickey-Fuller (ADF) test or differencing can be used to ensure this requirement is met.
   - Split data into an 80:20 ratio. This allows 80% of the data going towards the development/training of the model, with the remaining 20% of it going to testing/validation of the model. This split is important as it allows our model to be trained with an adequate amount of historical data, and be tested with recent data, simulating an environment of forecasting to see if the model performs as desired.

$$\boxed{R_t = \log P_t - \log P_{t-1} = \log\left(\frac{P_t}{P_{t-1}}\right)}$$

Figure 1. Log returns equation. Source: Jayati WALIA, "Returns", SimTrade, 2022

3. **Linear trend identification with ARIMA model**
   Linear trends in Bitcoin's price will be identified within the project using Box-Jenkins ARIMA. The steps to do this are:
   - Find the optimal parameters for p (autoregressive order), d (differencing order) and q (moving average order). This can be done using AIC/BIC.
   - Validation/examination must be done on values to attempt to uncover other residual trends still found within the data (Ljung-Box test).
   - Produce an ARIMA model capturing Bitcoin's short-term movements.

$$\underset{\text{Integrated}}{y_t^{(d)}} = c + \varepsilon_t + \underbrace{\phi_1 y_{t-1}^{(d)} + \phi_2 y_{t-2}^{(d)} + \ldots + \phi_p y_{t-p}^{(d)}}_{\text{Auto-Regressive}} + \underbrace{\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_p \varepsilon_{t-q}}_{\text{Moving Average}}$$

Figure 2. ARIMA model equation. Source: David Andres, "Introduction to ARIMA models, Machine Learning Pills, 2022

4. **Volatility clustering identification with GARCH model**

The GARCH model uses the residuals from the ARIMA model to identify volatility clustering. We do this by:

- Identify GARCH's required lag terms of p and q. This can be done using AIC/BIC.
- Fit lag terms using the ARIMA residuals.
- Produce our GARCH model for the volatility persistence of Bitcoin.

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \quad (1)$$

Where σ_t^2 is the conditional volatility, and ε_{t-1}^2 are squared unexpected returns for the previous period. ω would be positive always; and α and β would be non-negative (>=0). ε_{t-1}^2 are derived from a conditional mean equation that could be simple random walk model ($r_t = c + \varepsilon_t$), or AR (1) model ($r_t = c + \gamma * r_{t-1} + \varepsilon_t$), or any other ARMA model. But generally conditional mean equations are kept simple as it can cause convergence problems in GARCH estimation (Alexander, 2001). Where r_t is the returns from a financial series.

Figure 3. GARCH model equation. Source: Sarveshwar Kumar Inani, "GARCH Modelling", Sarveshwar Inani's Blog, 2016

5. **Simulation of hybrid ARIMA-GARCH model**

To create price predictions, we must do the following:

- Generate 1-step-ahead forecasts of both returns (conditional mean) and volatility (conditional variance).
- Next, incorporate these values within a Monte Carlo simulation to simulate thousands of possible price paths.
- From our simulations, we can obtain a Value-at-Risk (VaR) value as well as the worst-case scenarios (valuable insights for risk management).

$$\hat{y}_t = \boxed{\mu + \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{q} b_i \epsilon_{t-i}} + \boxed{\epsilon_t}$$

$$\epsilon_t = \sqrt{\sigma_t} z_t, \quad \sigma^2 = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2$$

Figure 4, ARIMA GARCH model equation. Source: Thomas Dierckx, "ARIMA-GARCH forecasting with Python", Medium, 2020

6. **Validating model performance**

To validate the performance of our hybrid model, we complete the following steps:

- Use metrics such as MSE, RMSE and MAE to evaluate forecasts for volatility compared to realised volatility.
- Compare our hybrid model compared to the standalone ARIMA and GARCH models to truly identify the impact of a hybrid model compared to using singular model. Do this using a Diebold-Mariano test.

Forecasting models	QLIKE	HMSE	HMAE
Mean	5.471***	4.457***	6.853***
Median	2.850***	6.154***	8.030***
Trimmed mean	2.747***	6.089***	7.548***
DMSPE(1)	6.146***	3.655***	6.758***
DMSPE(0.9)	6.091***	2.138**	3.623***

This table reports the Diebold-Mariano (DM) statistics for the paired comparison of the standard and iterated combination approaches. The DM statistic is calculated as

$$DM\ \ statistic = \frac{\overline{d}}{\sqrt{Var(d)}},$$

where $\overline{d} = \frac{1}{q}\sum_{t=m+1}^{m+q} d_t$, $d_t = L(\widehat{RV}_{c,t}, RV_t) - L(\widehat{RV}_{ic,t}, RV_t)$, $Var(d)$ is the variance of

Figure 5, Diebold Mariano model equation. Source: Yaojie Zhang, "Out-of-sample prediction of the oil futures market volatility, A comparison of new and traditional combination approaches", ResearchGate, 2018

7. **Interpreting and visualising results**

With all needed data gathered, we can:

- Create Q-Q plots to examine the model's expected outcome and evaluate the arrangement of residuals.
- Create a visualisation to compare the predicted price generated from our simulation against the actual price of Bitcoin, uncovering the true effectiveness of our model in predicting Bitcoin's price.

References:

BLOCKS, GX. 2022. "Bitcoin Characteristics." *Coinmonks*. March 11, 2022. https://medium.com/coinmonks/bitcoin-characteristics-8c9b2f1ae8.

Dierckx, Thomas. 2020. "ARIMA-GARCH Forecasting with Python." *Analytics Vidhya*. October 12, 2020. https://medium.com/analytics-vidhya/arima-garch-forecasting-with-python-7a3f797de3ff.

Inani, Sarveshwar Kumar. 2016. "Sarveshwar Inani's Blog: GARCH Modelling." *Sarveshwar Inani's Blog*. September 24, 2016. https://learningeconometrics.blogspot.com/2016/09/garch-modelling.html.

Andres, David. 2022. "Introduction to ARIMA Models - ML Pills." *ML Pills - Machine Learning Pills*. November 11, 2022. https://mlpills.dev/time-series/introduction-to-arima-models/.

Laura. 2021. "5 Features That Make Bitcoin a Unique Asset Class." *Argoblockchain.com*. Argo Blockchain. March 13, 2021. https://www.argoblockchain.com/articles/5-features-that-make-bitcoin-a-unique-asset-class.

Reiff, Nathan. 2018. "Why Bitcoin Price Predictions Are Unreliable." *Investopedia*. 2018. https://www.investopedia.com/tech/why-bitcoin-price-predictions-are-unreliable/#toc-bitcoin-price-predictions-are-hype-not-fundamentals.

Team, Towards AI. 2020. "Towards AI." *Towardsai.net*. December 23, 2020. https://towardsai.net/p/data-visualization/statistical-forecasting-of-time-series-data-part-4-forecasting-volatility-using-garch.

WALIA, Jayati. 2022. "Returns." *SimTrade Blog*. April 20, 2022. https://www.simtrade.fr/blog_simtrade/returns/.

Zhang, Yaojie, Feng Ma, and Yu Wei. 2018. "Out-of-Sample Prediction of the Oil Futures Market Volatility: A Comparison of New and Traditional..." *ResearchGate*, August. https://doi.org/10.13140/RG.2.2.35113.29282.

"Bitcoin Definition: How Does Bitcoin Work?" n.d. *Investopedia*. https://www.investopedia.com/terms/b/bitcoin.asp#toc-understanding-bitcoin.