Predicting Cryptocurrency Pricing

By: Siddharth Patel, Dylan Kim, Zihan Yu, Weijie Guan

DS 3000

Prof. Mohit Singhal

Research Questions/Motivation

Why study Cryptocurrency Volatility?

- Crypto markets are extremely volatile
- Understanding risk is crucial for investors
- Do traditional market metrics apply?

Questions of Interest:

- Are larger and established coins (like BTC and ETH) less volatile compared to that of smaller or emerging coins?
- What statistical signals explain short term volatility and how can we use these driven insights to predict how liquidity and market size relate to risk?

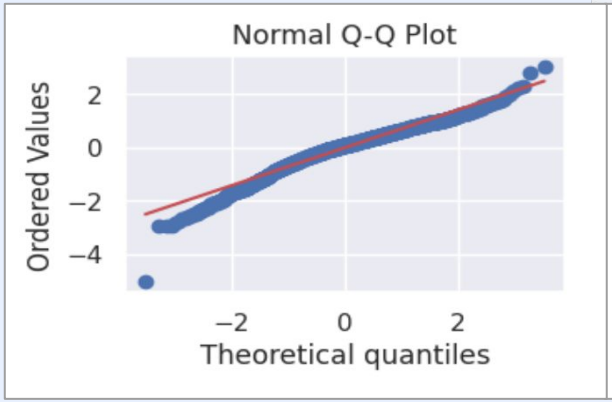
Data Source

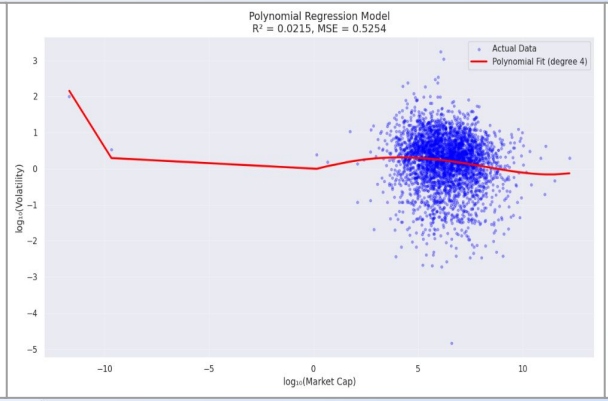
Our data is obtained from the ***CoinMarketCap API***.

The API provides *key data* such as price, market cap, and rank. These metrics are crucial in answering our questions of interest, providing insights on volatility and more.

Data Collection & Processing

Model 1: Linear Regression	Model 2: Polynomial Regression	Model 3: Added Int. terms & Dummy var.
Equation: $y = \beta_0 + \beta_1 x$ Tests: Does larger market cap → lower volatility? Where: x = log(cap) y = log(vol)	Equation: $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4$ Tests: Non-linear relationships Captures curvature in data	Model 2 + high_volume × (x, x^2, x^3, x^4) Tests: Does trading volume matter? high_volume = binary dummy (1 if >median)






```
high volume (above median): 1688 cryptocurrencies
low volume (below median): 1688 cryptocurrencies

Final feature matrix shape: (3376, 8)
First 5 rows:
[[1.22389329e+01 1.49791479e+02 1.83328787e+03 2.24374873e+04
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [1.15278644e+01 1.32891659e+02 1.53195703e+03 1.76601930e+04
  1.15278644e+01 1.32891659e+02 1.53195703e+03 1.76601930e+04]
 [1.10909589e+01 1.23009370e+02 1.36429188e+03 1.51313052e+04
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [1.10626765e+01 1.22382811e+02 1.35388145e+03 1.49775525e+04
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
 [1.08642961e+01 1.18032930e+02 1.28234470e+03 1.39317726e+04
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]]

Cross-Validation Results:
MSE for the interaction polynomial model = 0.524
R^2 for the interaction polynomial model = 0.0236
```

Machine Learning Models and Results

Model 1 — Worst

Linear Regression

MSE = 0.518
R^2 = 0.0144

Constant Variance and Normality violated
- Uneven spread + cluster w/ Residuals vs Fitted values plot
- Q-Q plot diverges from the line at both ends with heavier tails

Independence seemed okay, but not perfect
- Residuals vs Index plot shows random scatter with outliers

→

Model 2

Polynomial Regression

MSE = 0.525
R^2 = 0.0215

Improvement over Model 1
- Better R^2 indicates polynomial terms captured some non-linear relationships

Constant variance and Normality still violated

Independence maintained
- Residuals vs Index plots looks random

Residuals vs log Market Cap plot
- Scattered points, minimal pattern
- Confirmed weak but improved relationship

→

Model 3 — Best

Added Int. terms & Dummy var.

MSE = 0.524
R^2 = 0.0236

Converting continuous volume data to a binary dummy variable added predictive information
- Indicated by slightly increased R^2

Adding interaction terms and dummy variables slightly increased how much of a coin's volatility is explained by volume and market cap.

Conclusions

- Market fundamentals CANNOT predict cryptocurrency volatility ($R^2 < 0.0236$)

- Why did the models fail?
  - Crypto driven by sentiment, news, hype
  - Missing critical features (social media, regulation, macroeconomic factors)

- Future Research
  - Include sentiment analysis (Twitter/X or Reddit)
  - Time-series models
  - Specific models

- Takeaway: Crypto is fundamentally unpredictable using simple market metrics