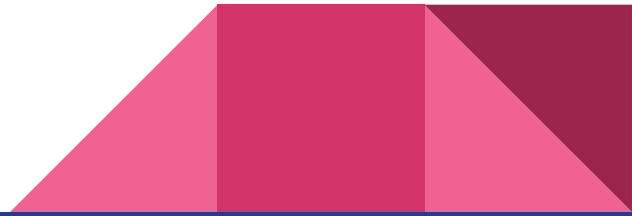
# Can we find a relationship between the stock prices of selected companies, gold prices and the bitcoin price?

Kareem Ali, Josh Green, Kyrian Puttileihalat, Mark Haakman

# Contents





# Data gathering/handling

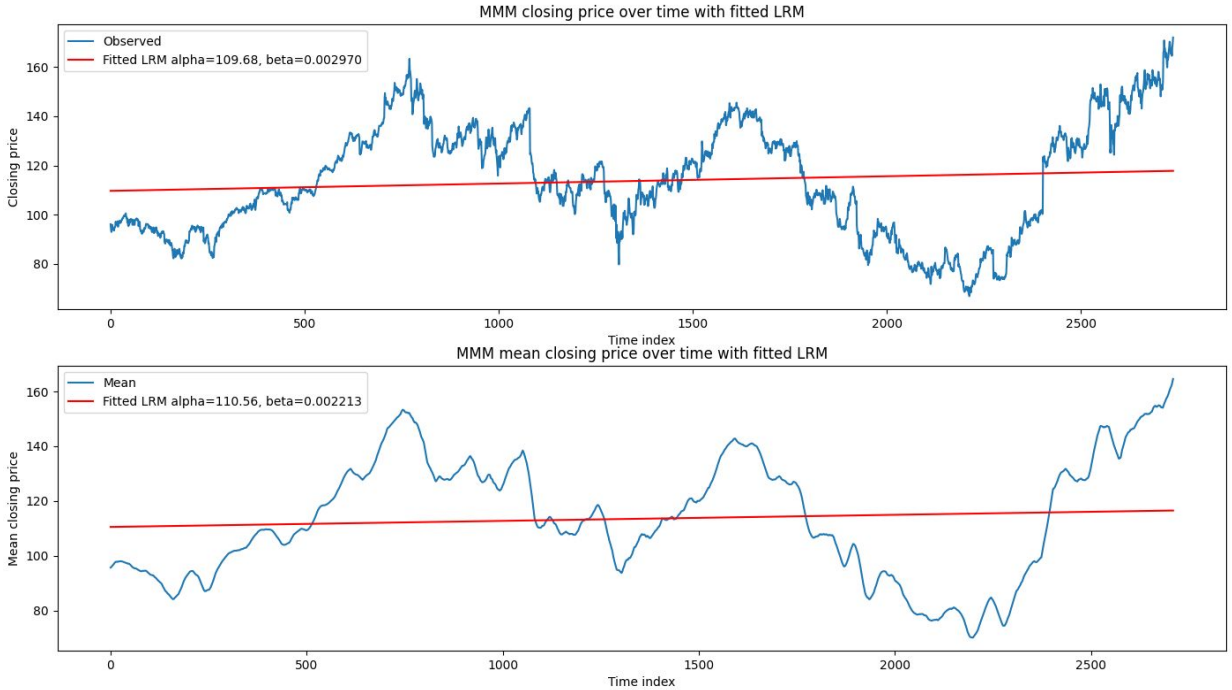
- Financial Market group
- Time-series
- 30 John Dow companies, Gold prices, Bitcoin prices
- Kaggle datasets
- Incomplete sets
- Python scraper

# R0: Stationary in mean (setup)

- Are all prices stationary in mean?
  - H0: All prices are stationary in mean (beta=0)
  - H1: All prices are not stationary in mean (beta != 0)
- Testing for slope by fitting LRM
- Rolling mean
- Bootstrapping
- Parameters/methods
  - LRM method
  - Bootstrap method
  - Window size

$$
\begin{aligned}
& X\hat{\beta} = y \\
\iff & X^{\mathrm{T}} X\hat{\beta} = X^{\mathrm{T}} y \\
\iff & \hat{\beta} = \left(X^{\mathrm{T}} X\right)^{-1} X^{\mathrm{T}} y.
\end{aligned}
$$

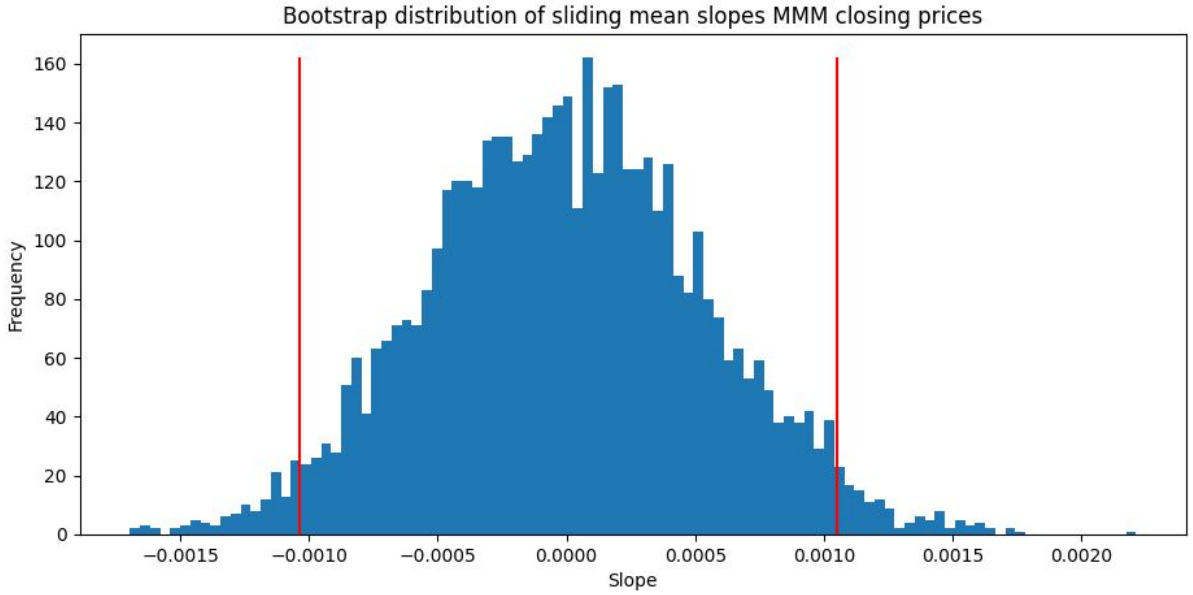
# R0: Stationary in mean (adj closing price)



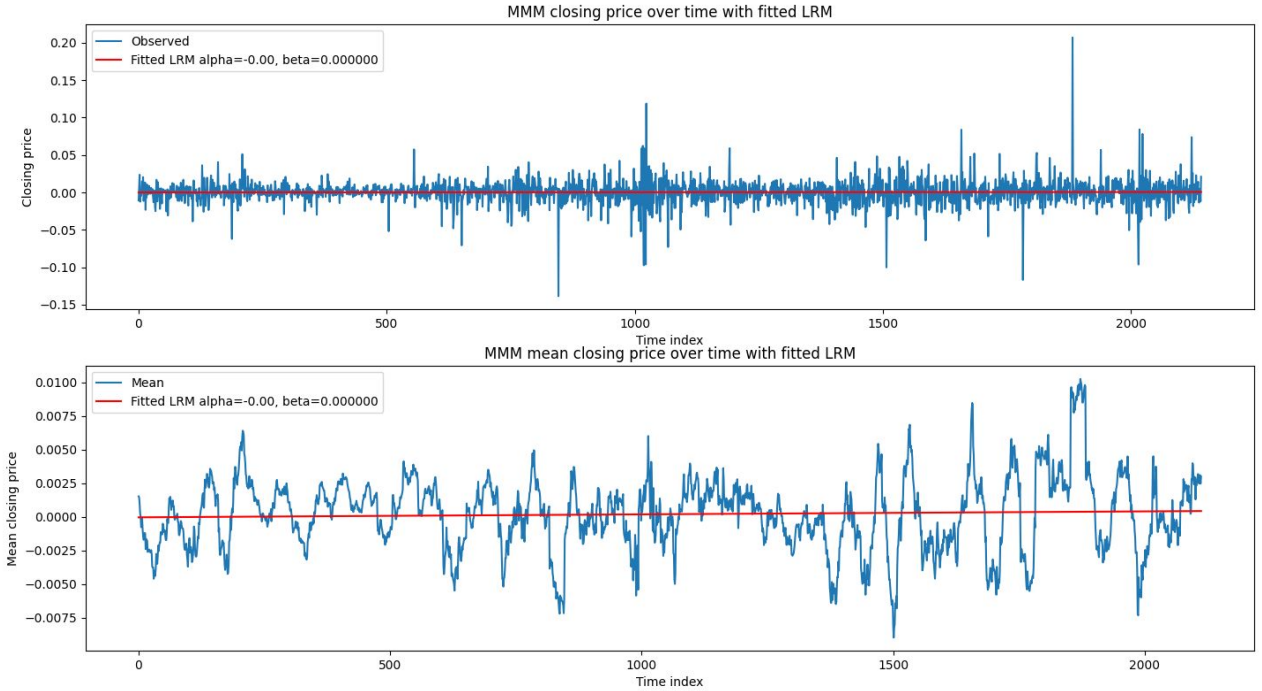
# R0: Stationary in mean

Observed slope: $\beta$=0.002213

CI bootstrap: [-0.0010, 0.0010]



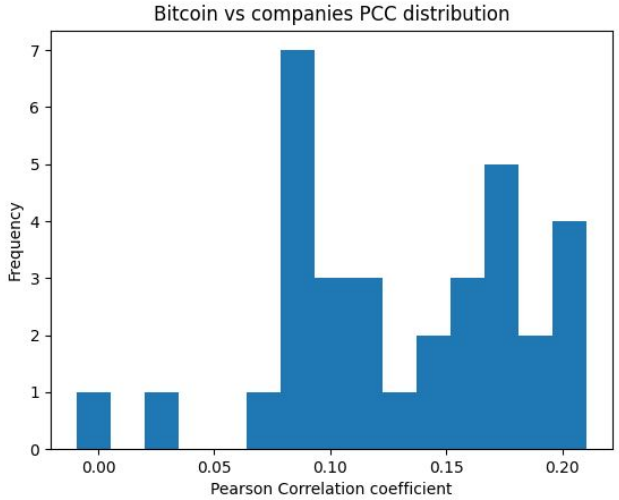
# R0: Stationary in mean (log of adj closing price)

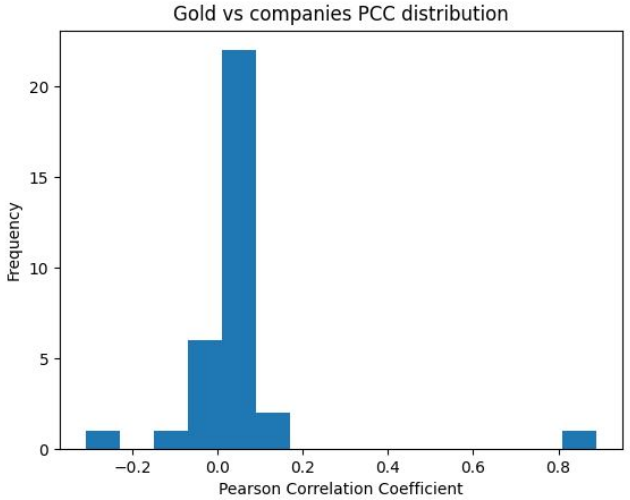


# R1-3: (Delayed) Correlation

- Pearson Correlation Coefficient
- Testing over delays
- GC vs GC=F correlation

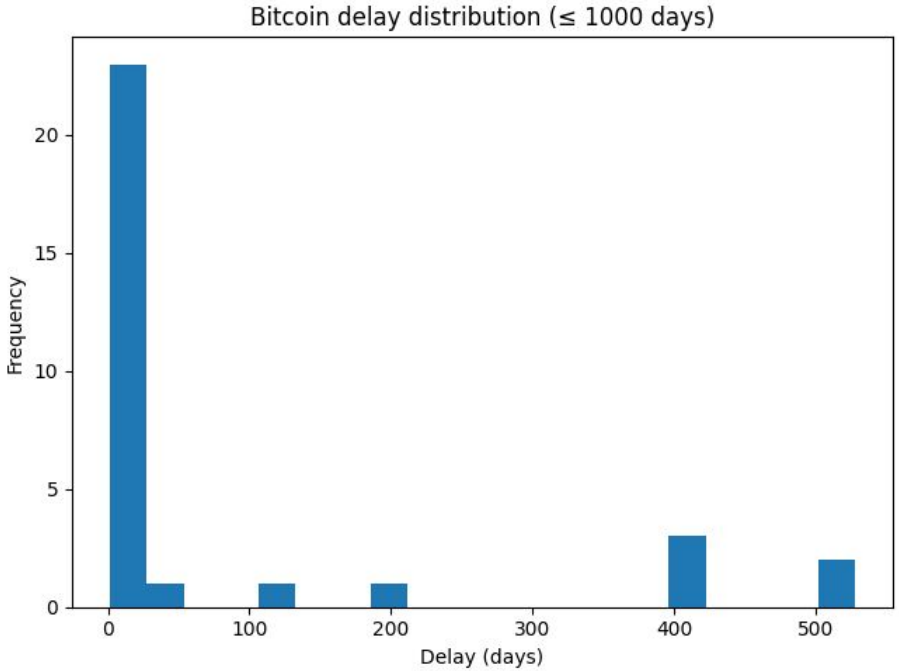
$$r = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum(X_i - \overline{X})^2 \sum(Y_i - \overline{Y})^2}}$$

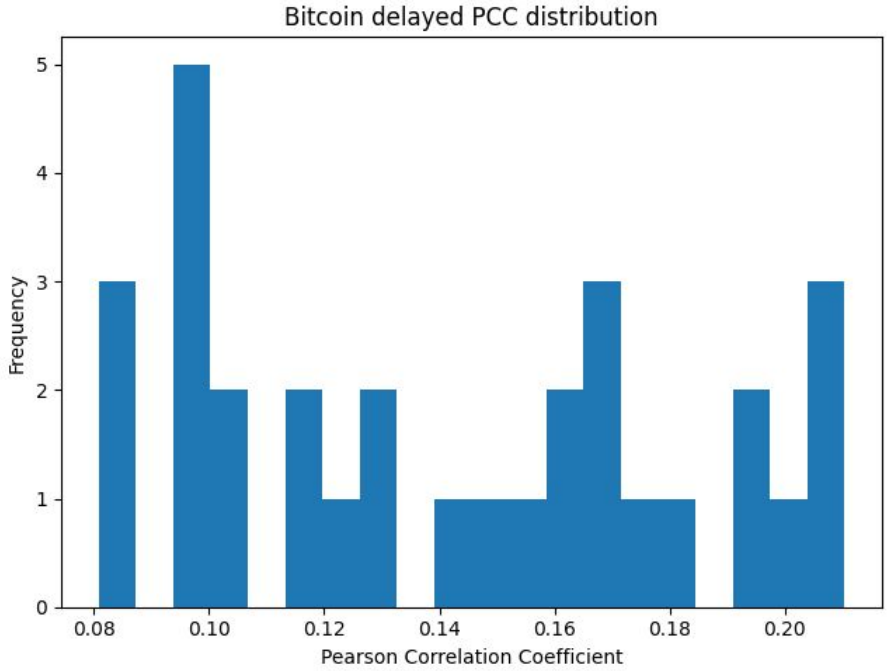


# R1-3: (Delayed) Correlation (bitcoin delay)

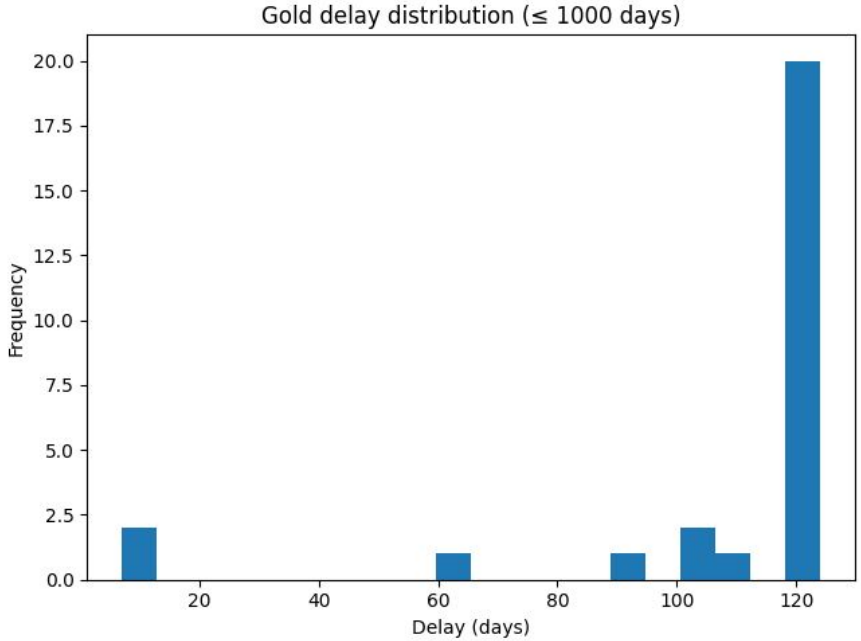
Filtering out any delays over 1000 days

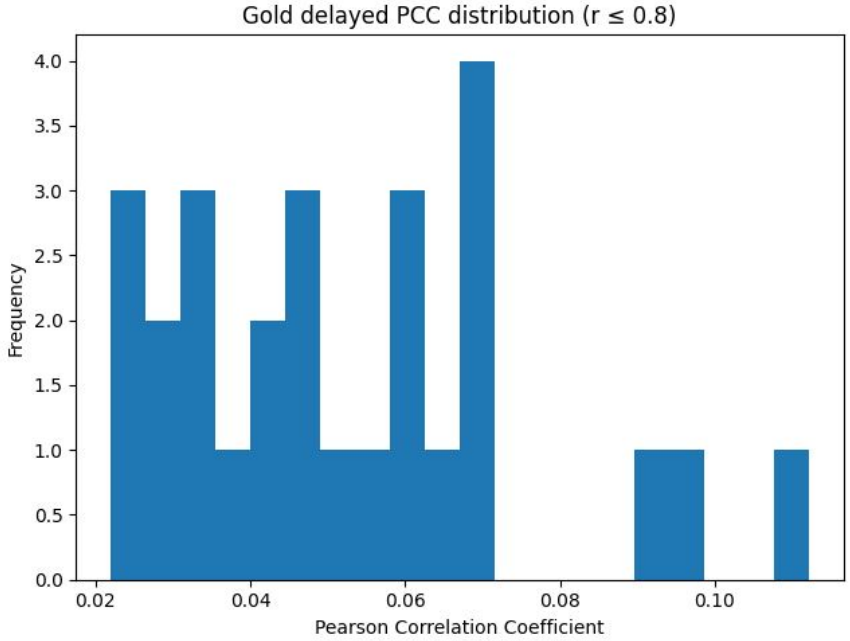


# R1-3: (Delayed) Correlation (gold delay)

Filtering out any delays over 1000 days and PCC value > 0.8 (GC vs GC=F)

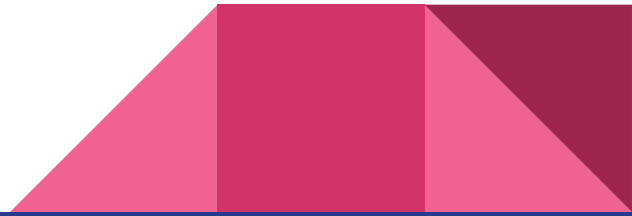


# R1-3: (Delayed) Correlation

Key observations:

- Obvious correlation between GC and GC=F
- Bitcoin has low delay and higher correlation
- Gold has higher delay and lower correlation



# R4: Trading Volume vs Stock Returns (Setup)

**Does trading volume explain same-day returns?**

**H0:** Trading volume has no effect on stock returns (beta = 0)
**H1:** Trading volume affects stock returns (beta != 0)

**Data transformations**

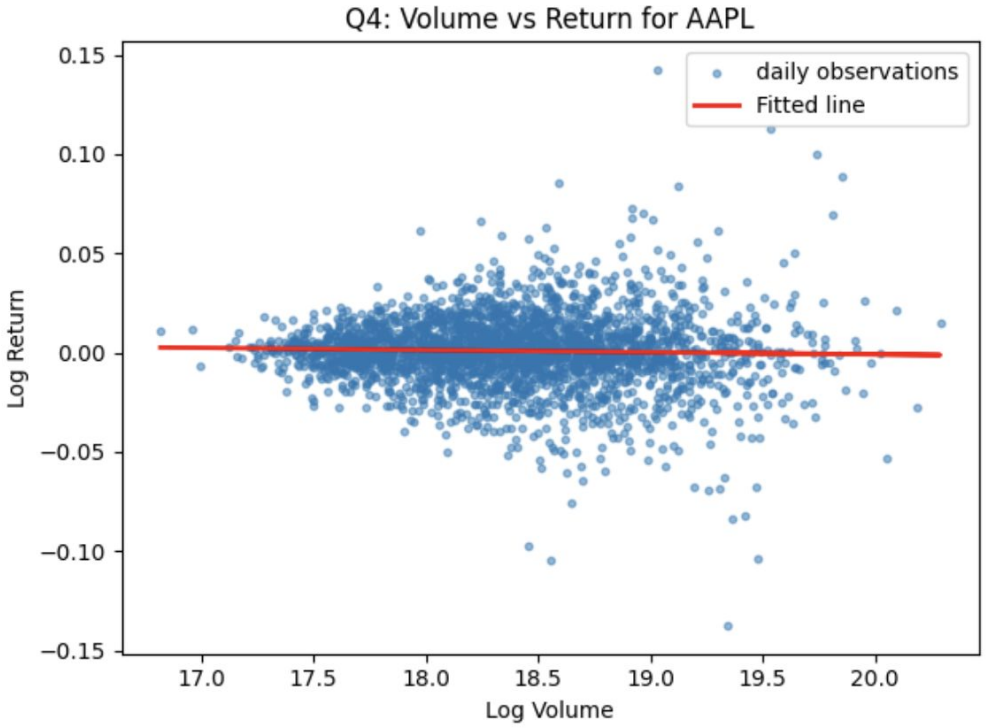
- Prices are transformed into daily log returns to make them stationary
  - $r_t = \log(P_t) - \log(P_{t-1})$

- Trading volume is transformed to log to remove skewness and make it readable
  - $x_t = \log(V_t)$

**Testing for slope by fitting LRM separately for each stock**

**Compute T-tests**

# R4: Trading Volume vs Stock Returns



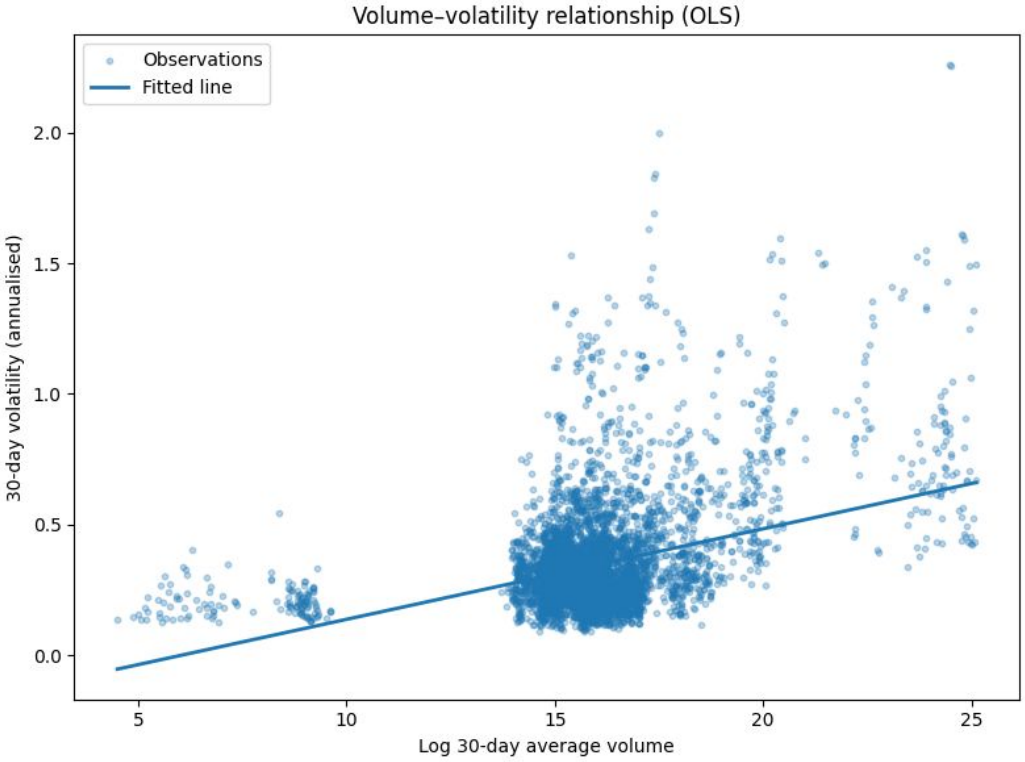

Effects are statistically weak and economically negligible

| | Ticker | Slope (log Volume → Return) | t-stat |
|---|---|---|---|
| 0 | AAPL | −0.001096 | −1.7 |
| 1 | AMGN | −0.001413 | −1.84 |
| 2 | AMZN | −0.0004 | −0.47 |
| 3 | AXP | −0.00205 | −2.59 |
| 4 | BA | −0.00142 | −2.22 |
| 5 | BTC-USD | −0.000106 | −0.33 |
| 6 | CAT | −0.001788 | −2.39 |
| 7 | CRM | −0.001299 | −1.52 |
| 8 | CSCO | −0.002375 | −2.91 |
| 9 | CVX | −0.003002 | −3.63 |
| 10 | DIS | −0.000441 | −0.61 |
| 11 | GC=F | nan | nan |
| 12 | GLD | 0.000781 | 2.1 |
| 13 | GS | −0.00135 | −1.64 |
| 14 | HD | −0.00278 | −3.69 |
| 15 | HON | −0.002111 | −3.01 |
| 16 | IBM | −0.002302 | −3.3 |
| 17 | JNJ | −0.000493 | −0.88 |
| 18 | JPM | −0.002918 | −3.76 |
| 19 | KO | −0.001334 | −2.32 |
| 20 | MCD | −0.001134 | −2.14 |
| 21 | MMM | −0.000828 | −1.28 |
| 22 | MRK | −0.000984 | −1.45 |
| 23 | MSFT | −0.001568 | −1.87 |
| 24 | NKE | −0.002305 | −2.93 |
| 25 | NVDA | 0.00017 | 0.15 |
| 26 | PG | −0.001001 | −1.76 |
| 27 | SHW | −0.000919 | −1.3 |
| 28 | TRV | −0.001701 | −2.23 |
| 29 | UNH | −0.002507 | −3.8 |
| 30 | V | −0.002616 | −3.61 |
| 31 | VZ | −0.001741 | −2.88 |
| 32 | WMT | −0.000992 | −1.64 |
| 33 | ^IRX | nan | nan |
| 34 | ^TNX | nan | nan |

# R5: Does the volume affect the stock's volatility?

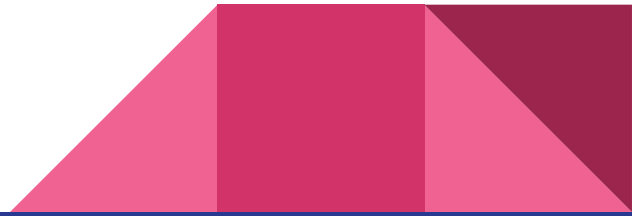
(script output)

```
Tickers used: ['AAPL', 'AMGN', 'AMZN', 'AXP', 'BA',
'BTC-USD', 'CAT', 'CRM', 'CSCO', 'CVX', 'DIS', 'GC=F',
'GLD', 'GS', 'HD', 'HON', 'IBM', 'JNJ', 'JPM', 'KO',
'MCD', 'MMM', 'MRK', 'MSFT', 'NKE', 'NVDA', 'PG', 'SHW',
'TRV', 'UNH', 'V', 'VZ', 'WMT', '^IRX', '^TNX']
Number of observations in panel: 69696
        Date ticker   vol_30d  log_vol_avg_30d
0 2015-03-03   AAPL  0.433054        19.343366
1 2015-03-03   AMGN  0.418438        15.081996
2 2015-03-03   AMZN  0.574541        18.348809
3 2015-03-03    AXP  0.421993        16.108912
4 2015-03-03     BA  0.357614        15.553565

Regression results (using autoregression model):
alpha = -0.209631 (SE = 0.005862, t = -35.76)
beta  = 0.034658 (SE = 0.000364, t = 95.09)
sigma = 0.215343
R^2   = 0.164
n = 69696

Interpretation of beta (semi-log effect sizes):
If 30-day avg volume is 1% higher, predicted vol_30d
increases by 0.000177.
If 30-day avg volume is 10% higher, predicted vol_30d
increases by 0.001694.
If 30-day avg volume is 50% higher, predicted vol_30d
increases by 0.007209.
If 30-day avg volume is multiplied by 2x, predicted
vol_30d increases by 0.012323.
```

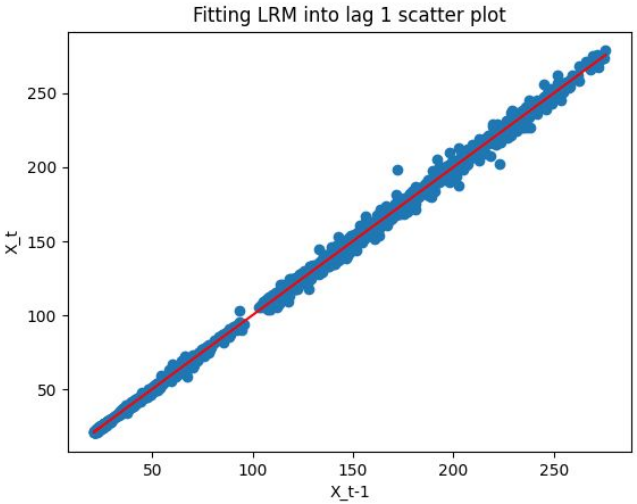


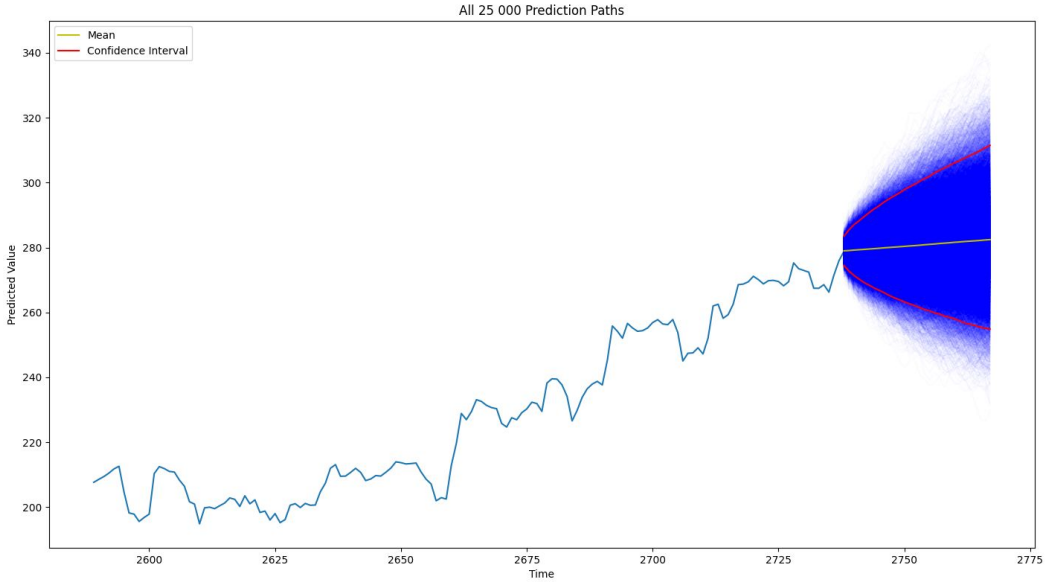
# R6: Predicting the Stock price

- AR(1) model
  - Pt+1 = alpha + beta * Pt + epsilon_t
- Time as variable
- LRM model fitting to X_(t-1) vs X_t
- Parameters
  - alpha
  - beta
  - sigma
  - # parameter fits
  - # prediction paths



# R6: Predicting the Stock price  (resulting plots)



# Conclusion

- Prices not stationary, log prices seem to be stationary
- Bitcoin has a stronger correlation with less delay compared to gold
- Trading volume has very little explanatory power for stock returns
- In contrast, a higher 30 day average volume is linked to higher 30 day annualised volatility
- AR(1) predicting method is a basic assessment of how predictable price movements are under linear autoregressive assumptions
- Overall, the results suggest that market activity influences how much prices move and how assets move together, but not the direction of price changes.

# Discussion and Limitations

- Sliding mean is not a method for concluding stationarity
  - consider periodic functions
- Filtering with correlations is an arguable choice with few datasets
- AR(1) is a simple benchmark model (only previous stock)
  - consider higher lag or models such as ARMA
- Volume is more informative for volatility and not for direction of price
- Volume effects might have a delay
  - consider using lagged volume for a better test on returns/volatility
- Form misspecification
  - consider using non-linear models if they fit the data better.
  - could find complex/non-linear effects