Cryptotoken Statistics At Intraday Frequencies

Brian K. Boonstra

5/17/2019

Conclusions

Basic Findings

- Cryptotoken returns (in USD terms) are fundamentally fatter-tailed than equity returns
- The fatter tails are *not* explained by (slow-moving) stochastic volatility
- Bitcoin returns are fatter-tailed than Ethereum returns
- This is true on both daily and high-frequency time scales
- Combined with tiny spreads, this means classic market-making is unprofitable

So many presentations rush to get to the conclusion.

I will give the conclusion first, so you needn't pretend to be interested for the rest.

How Did We Get Here?

- Get high-frequency (and daily) cryptotoken market data
- Download similar data for some other assets, mainly US equities, for comparison
- Run some simple statistics, via some not-so-simple coding

Market Data

Cryptotoken Markets

Some cryptotokens are widely, and liquidly, traded

- Bitcoin (BTC) daily volumes: ~$200MM (US Dollars)
- Ethereum (ETH) volumes: ~$40MM
- These volumes are totaled only from credible exchanges (Matt Hougan, Bitwise Asset Management, March 2019)
- Cryptotoken markets are roughly as fragmented as US equity markets, though without Reg NMS

Prominent Example: Coinbase (GDAX)

Coinbase has several features making it attractive for study

- US-licensed
- Has engaged regulators
- High volume
- High-frequency data feed available to all users
- Enjoys partnership with traditional financial institutions

Data Used In This Study

- Full Level 3 (all orders)
- Three pairs: BTC/USD, ETH/USD and BTC/ETH
- 60 days of 2019 data, largely contiguous
- Roughly 4MM messages/day (300MB/day, compressed)

Analysis Plans

What Are Returns?

- Classic definition

$$r = \log\left(\frac{P_n}{P_{n-1}}\right)$$

- HF definition (Johnson, 2010) to mitigate discrete tick sensitivity…

Defining High Frequency Returns

- Start with EWMA at characteristic time c

$$p_t^{(c)} = \mathrm{EWMA}\left(\{P_s : s \leq t\}; \lambda = c^{-1}\right)$$

- Define return relative to weighted average of past prices around time $t - c$

$$r = \log\left(\frac{P_t}{p_t^{(c)}}\right)$$

- Or, just as usefully

$$r = \frac{P_t}{p_t^{(c)}} - 1$$

Time-Based EWMAs

Because fast markets result in numerous updates within a short time span, it is important for our EWMA to be time-based, rather than count-based.

$$p_{\text{new}} = e^{-\lambda \Delta t} p_{\text{old}} + (1 - e^{-\lambda \Delta t}) P_{\text{new}}$$

Time-Based Volatility

Volatility is change over time, so we must be careful when forming EWM volatility. The formulas are tricky.

Rather than using complex formulas, it's simpler to take advantage of the fact that we have historical data here, and convert:

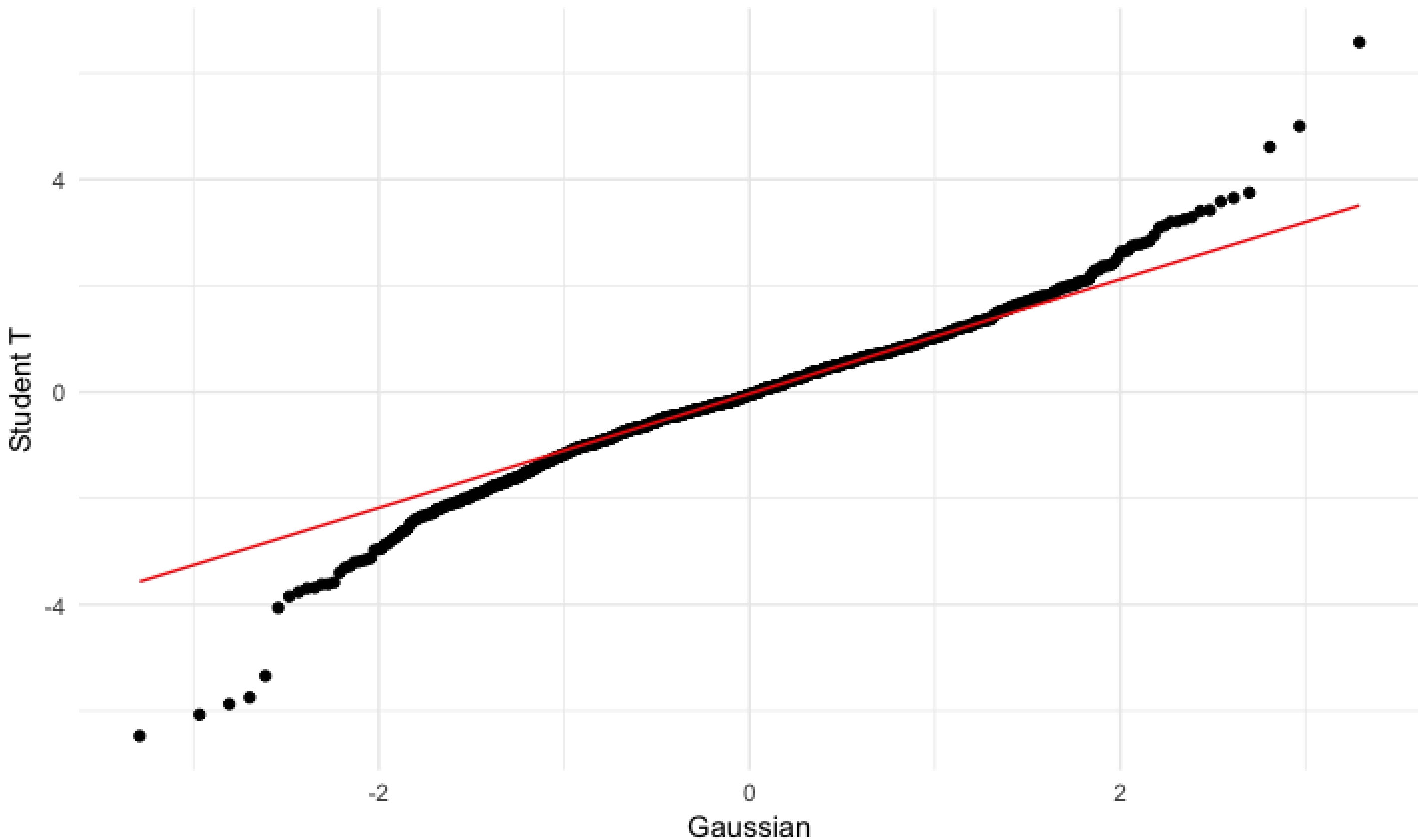
- Downsample to regular time intervals
- Compute an EWMA of $(r - r_{\text{avg}})^2$, or even just r^2

Shape Of Return Distributions

We are interested in the distribution of returns for cryptotokens.

- Do not care too much about location and scale
- Interested in shape, especially of tails
- Histograms are a familiar approach
- Better approach: QQ plots (just as we get from `plot.lm`)

Concept: Quantile-Quantile Plot



Fat tails are deviations from the line, Student T versus the gaussian

What Is The Price?

For equities, standard analysis at daily or lower frequencies typically uses closing prices. That doesn't work at higher frequencies.

We can use mid price between bid B and offer A

$$P_{\text{mid}} = \frac{1}{2}(B + A)$$

Central Price

Alternatively, we can use a central price that is aware of the number of shares S_B, S_A presently available at the bid or offer (Lipton 2014).

We weight the two prices by the square root of opposite share count

$$P_{\text{central}} = \frac{1}{\sqrt{S_B} + \sqrt{S_A}} \left(B\sqrt{S_A} + A\sqrt{S_B} \right)$$

Return Sampling

For some purposes, such as system design, using returns generated at the irregular high-frequency timestamps is appropriate.

More generally, we desire to do our statistics on a time-weighted subsample of the full high frequency returns.

If we are interested in timescales of length Δt, then we oversample a bit, forming a regular series with samples at intervals of $\Delta t/K$, for some $K > 1$.

Technology Comments

Timestamps

UTC timestamps at high-frequency resolutions are problematic at double precision

- Best addressed by `nanotime` and `bit64`
- Need to be careful with maps to double precision when time intervals are needed

As with so many presentations, this one is a thinly disguised advertisement for packages developed by R in Finance organizers.

High Frequency Analysis Technology

- At these data densities, it is important avoid loops and especially $O(N^{2+})$ algorithms
  - Loops often creep in when you use someone's package for a computation
  - Safer to have certainty with a little `C++` code
- Time-based EWMAs are not generally present in the common open-source packages, but are trivial to implement at high speed using `Rcpp`

Coinbase Data Feed

Silicon Valley, rather than finance industry, pedigree exhibited by message formats

- Coinbase:

```
"{'type': 'open', 'side': 'sell', 'price': '0.03743000',
'order_id': '88d50b31-d5c6-445b-9134-05989d05e165',
'remaining_size': '0.01000000', 'product_id': 'ETH-BTC',
'sequence': 1601762096, 'time': '2019-02-18T19:42:34.700000Z'}"
```

- FIX (CME):

```
'1128=9\x019=265\x0135=X\x0134=10065111\x
0152=20141013152659076\x
0175=20141013\x01268=2\x01279=2\x0122=8\x0148=656784\x0183=215750\x
01279=0\x0122=8\x0148=656784\x0183=215751\x01269=1\x01270=1850\x
01271=291\x01273=152659000\x01336=2\x01346=6\x
011023=3\x0110=008\x01'
```

Data Gaps

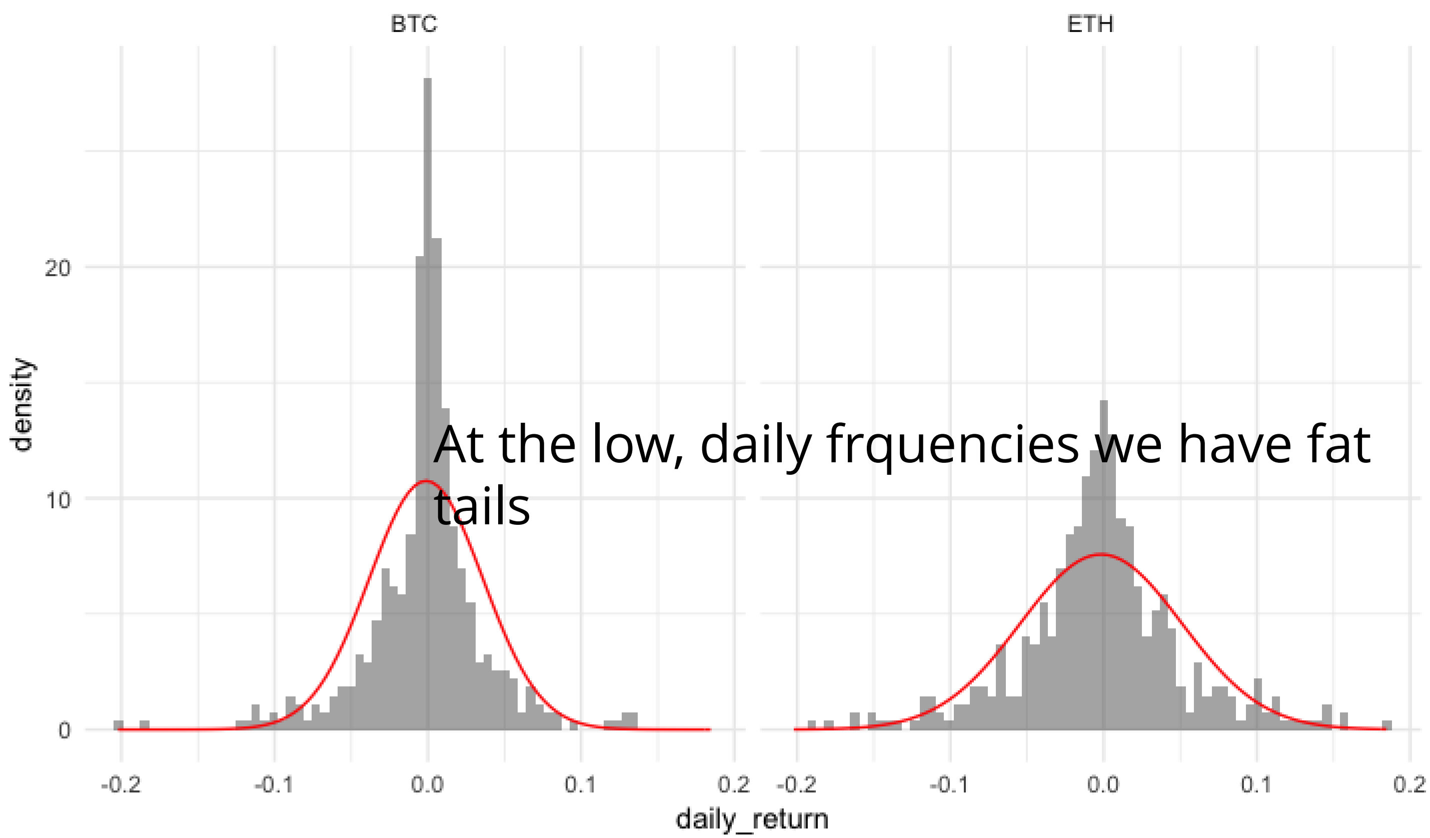
Our data has many gaps, which generally can be due to exchange hours, trading halts and system hiccups.

A time-based EWMA naturally resets these, while count-based EWMA implementations can get in trouble.

We have to be sure to invalidate our volatilities in places where gaps occur.

Statistical Observations

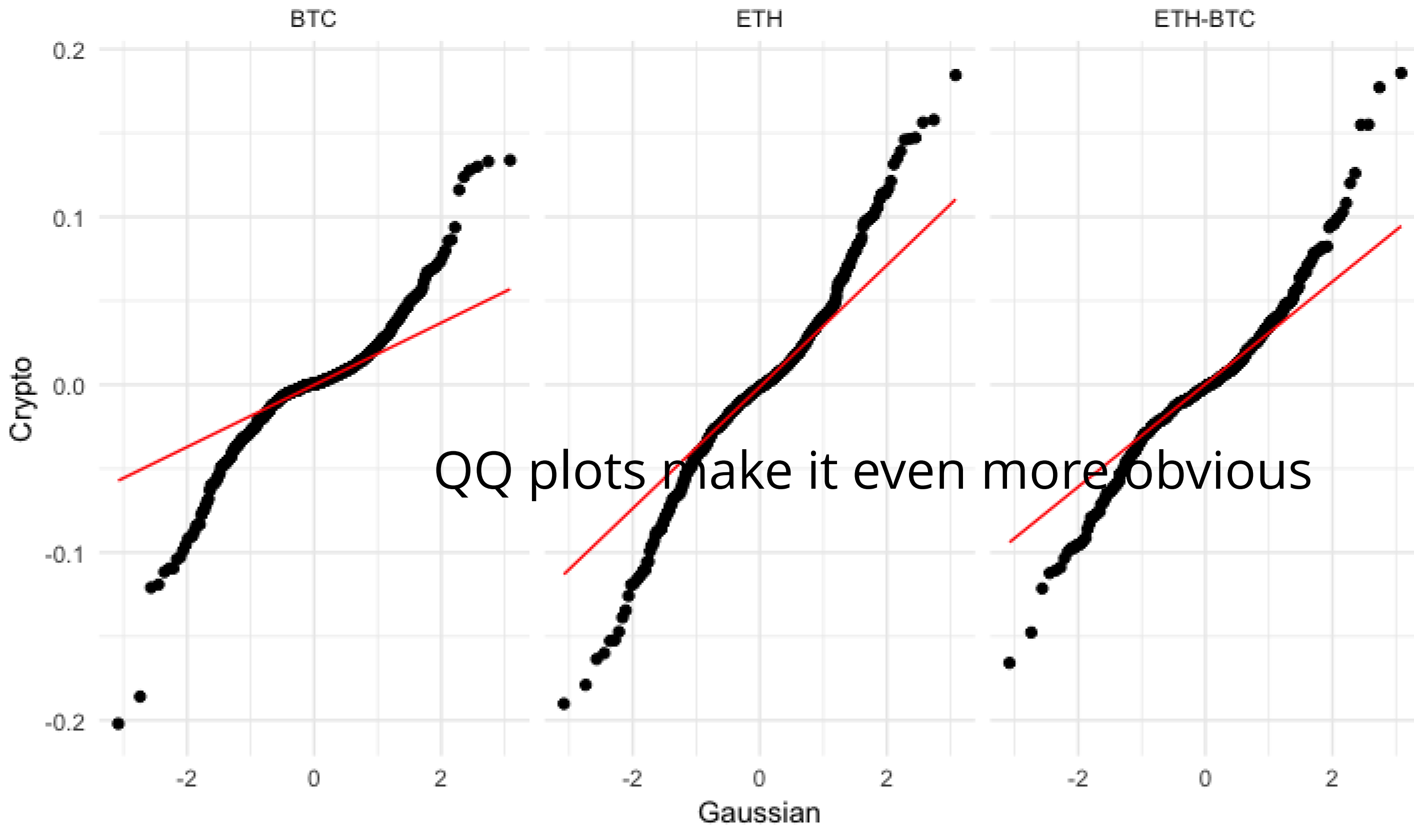
Daily Return Histograms



At the low, daily frquencies we have fat tails

Returns appear leptokurtotic

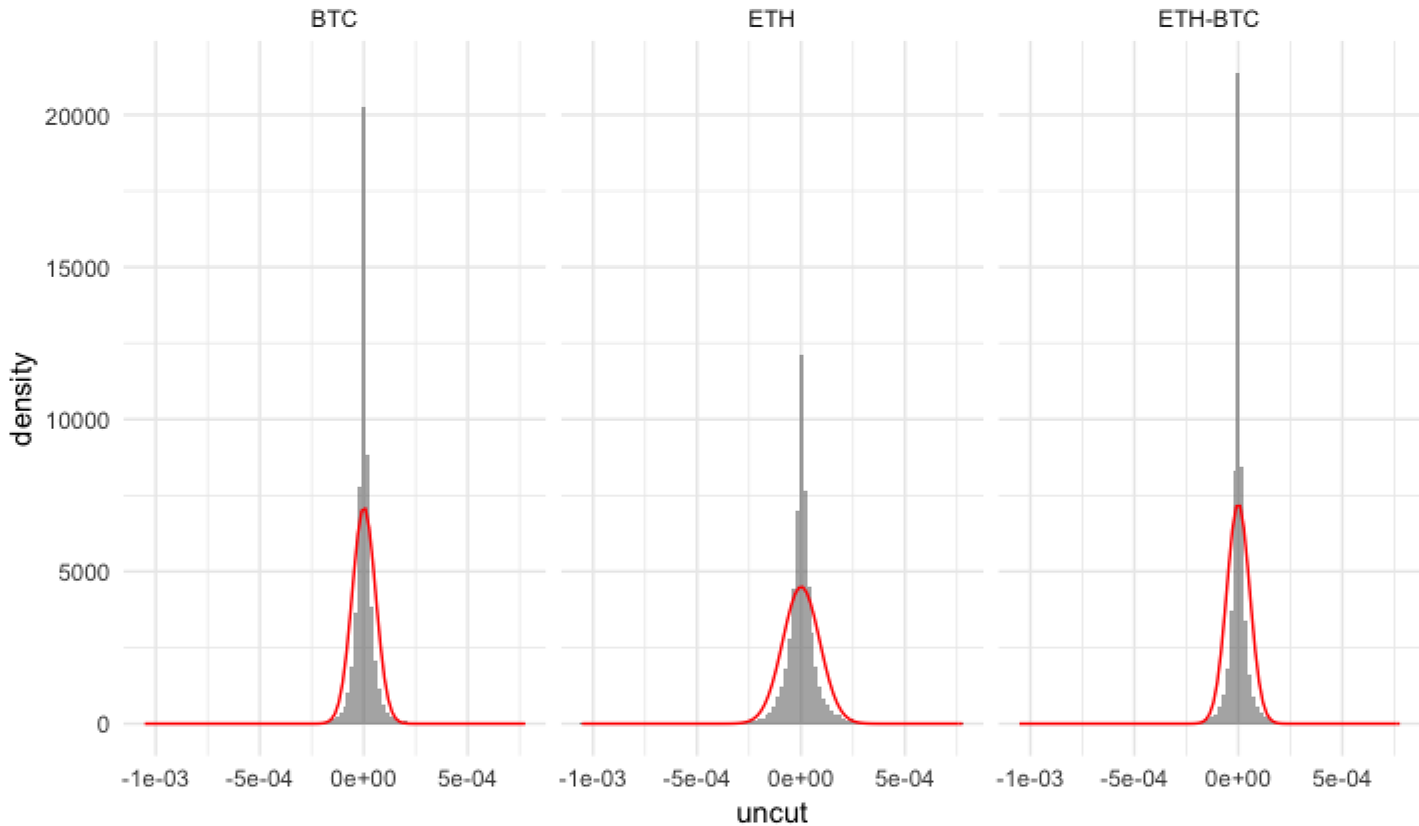
QQ Daily Crypto Returns



QQ plots make it even more obvious

The cryptotoken fat tails are highly visible

HF Return Histograms



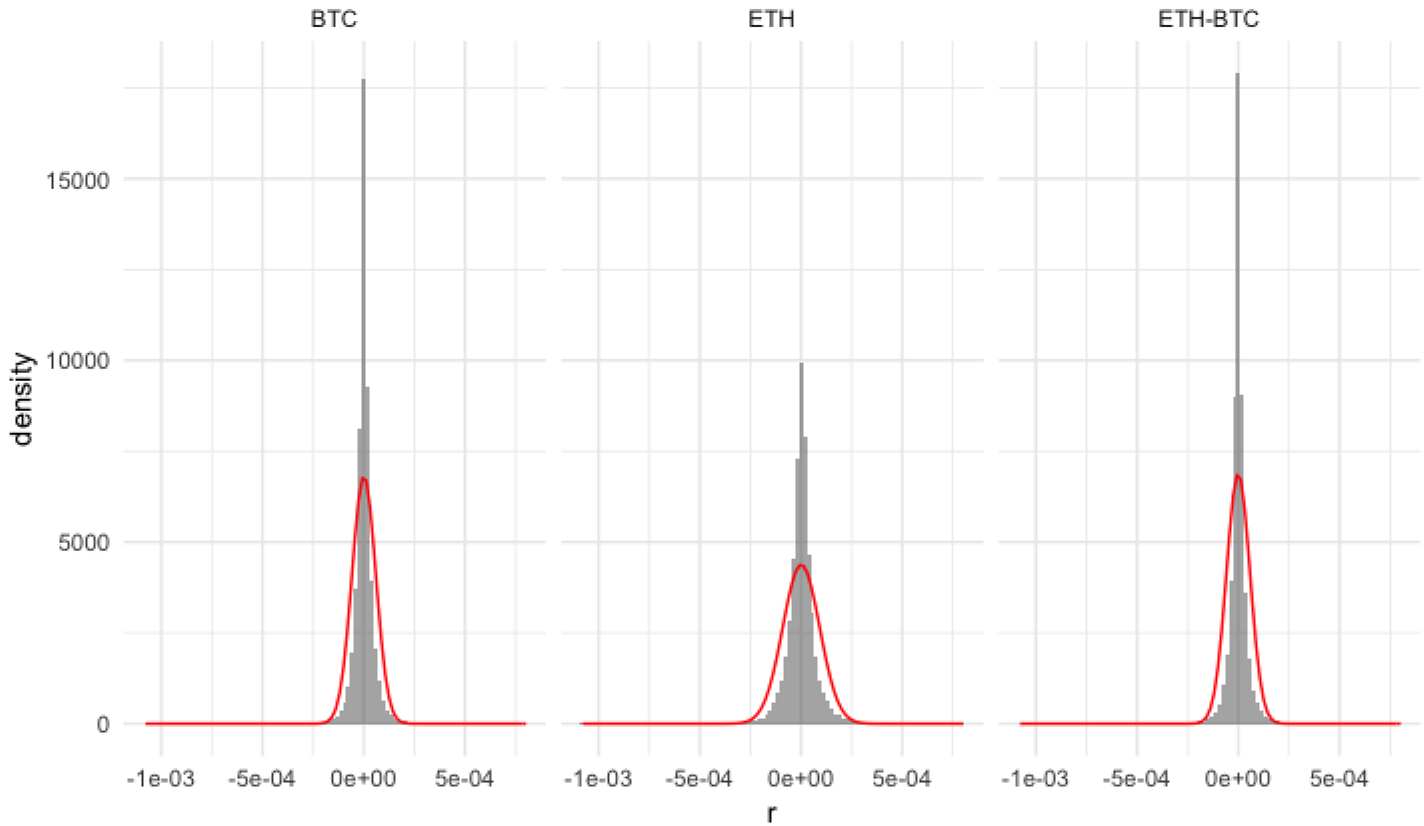
High-frequency returns are dominated by zero

Removing Zero

Since these are dominated by zero, it is often more informative to examine only significant values in the distribution

$$\{\hat{r}\} := \left\{ r : |r| > \frac{1}{20} \mathrm{Median}\left(\{r\}\right) \right\}$$

Modified Histograms



Removing zero did not make much qualitative difference

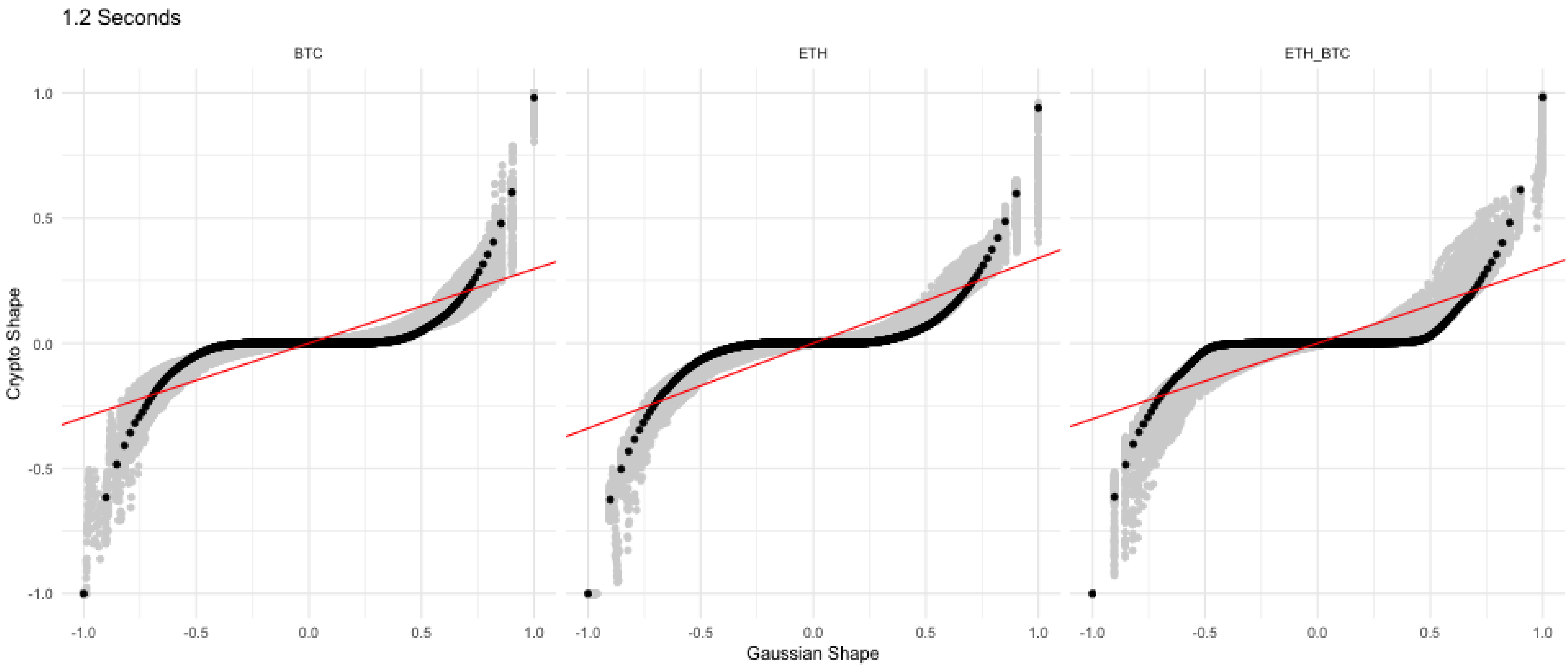
The Distribution Shapes At High Frequency

QQ At High-Frequency Time Scales

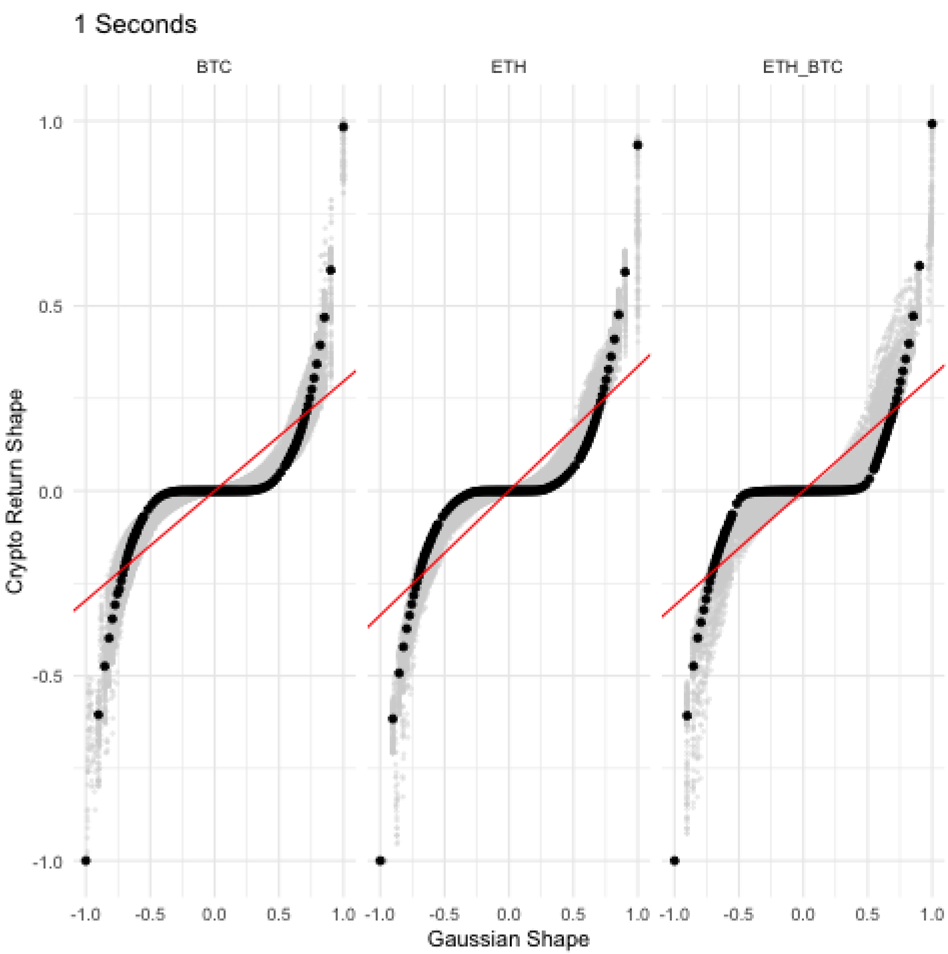
Time Scale:

5 200

1 21 41 61 81 101 121 161 200



QQ By Time Scale (Cycling)



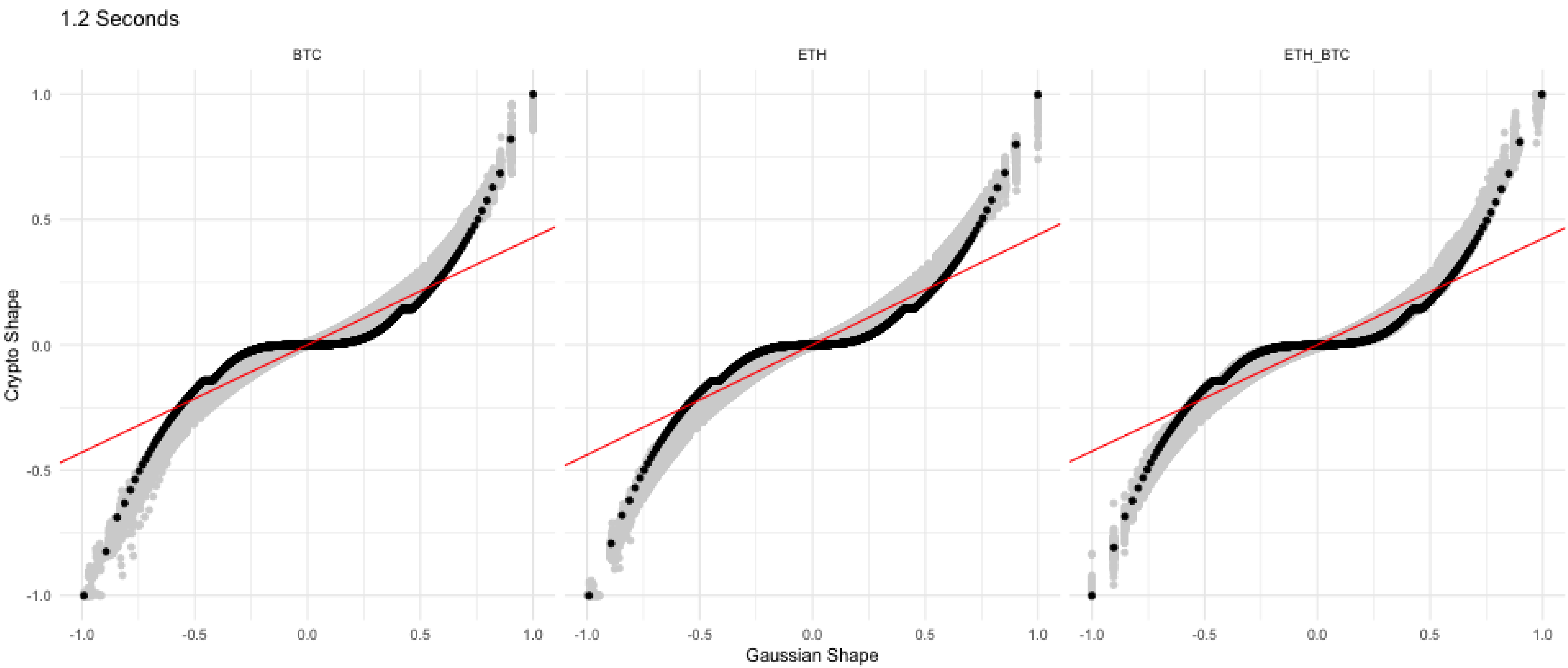
BTC QQ by time scale

Vol Normalized HF QQ

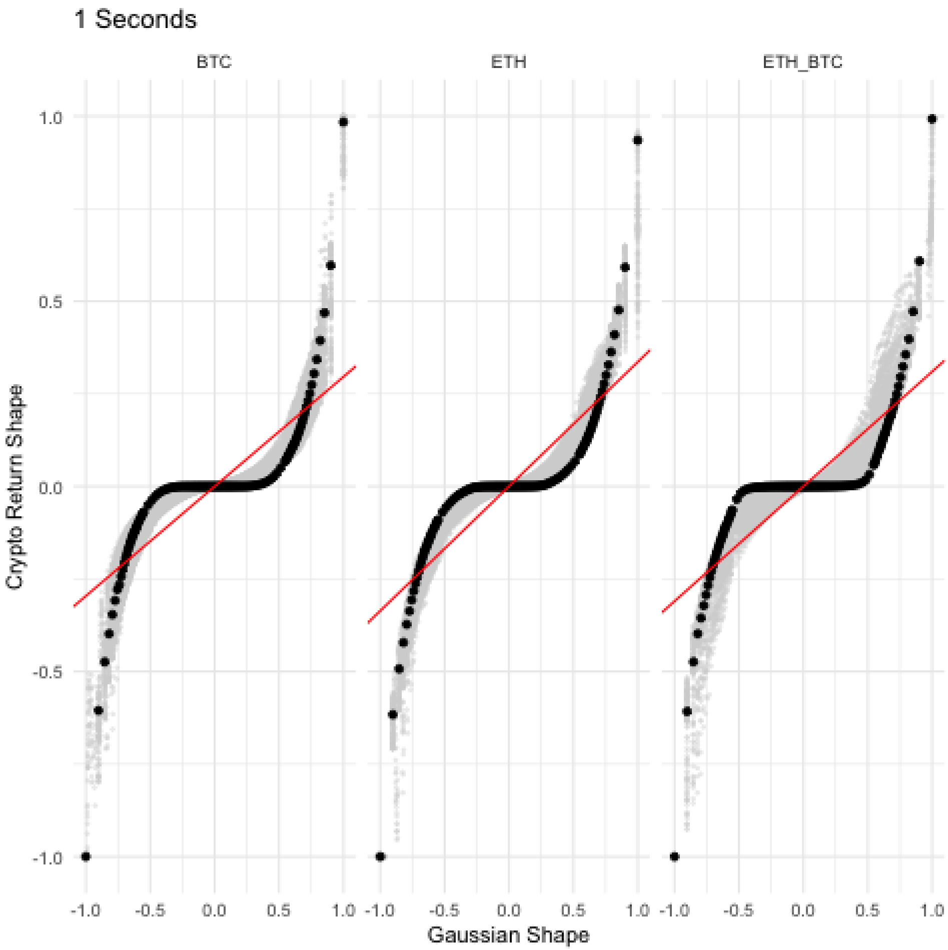
Time Scale:

5 200

1 21 41 61 81 101 121 161 200



Vol Normalized (Cycling)



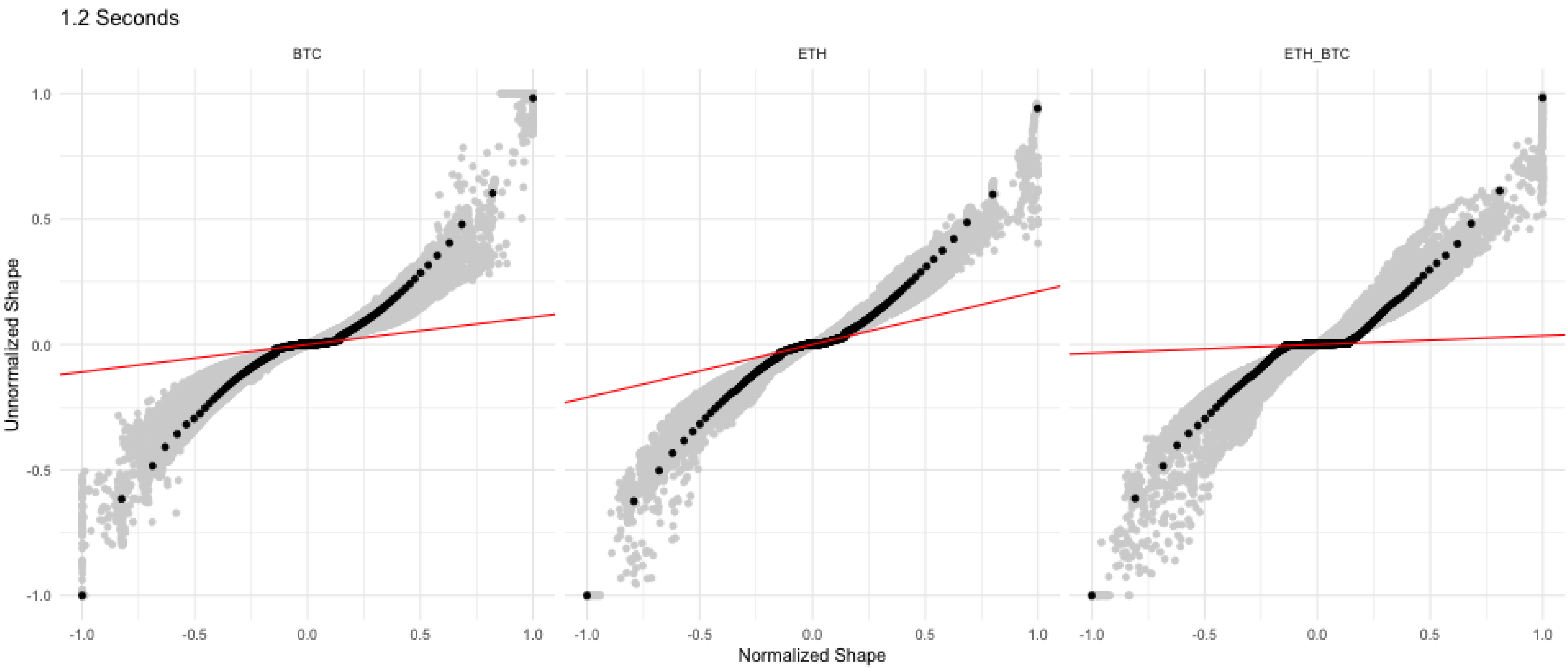
Normalized return distributions are still leptokurtotic

How Much Difference Did Normalization Make?

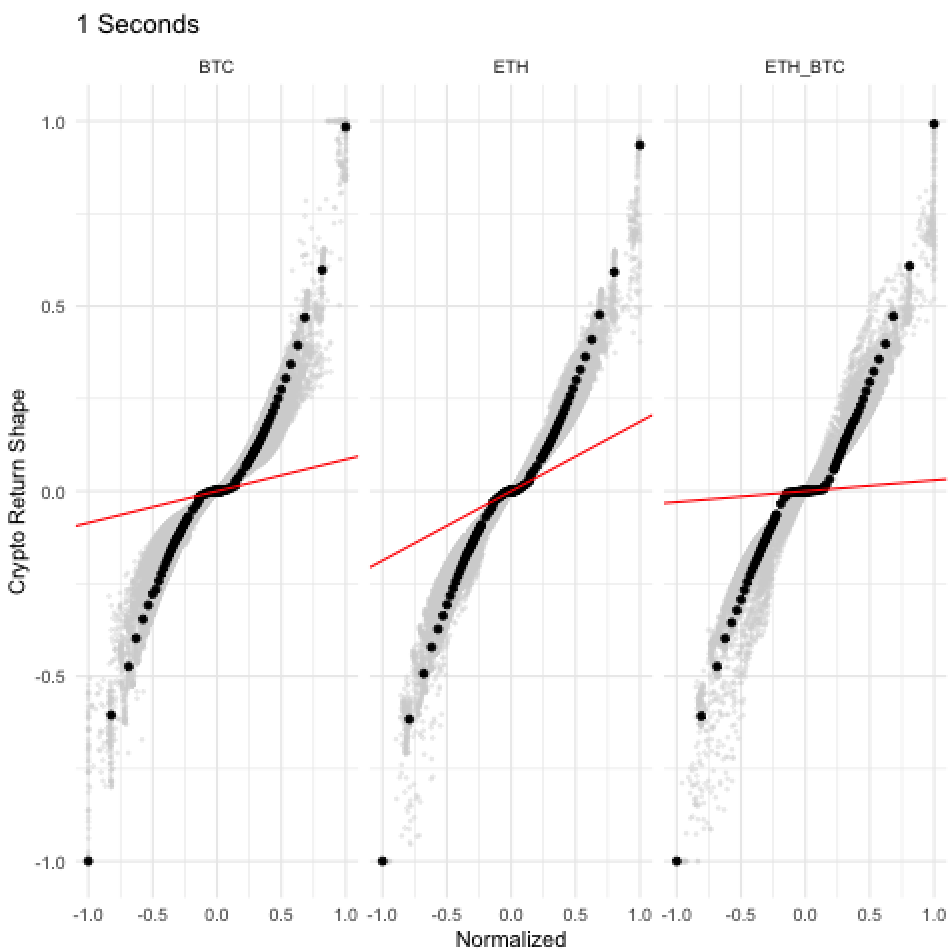
Time Scale:

5 200

1 21 41 61 81 101 121 161 200



Normalization Difference (Cycling)



Normalization did partially remove fat tails

Do BTC and ETH Differ?

Examining QQ plots side by side is one way to decide, but we have a more powerful tool: making a QQ plot of one versus the other.

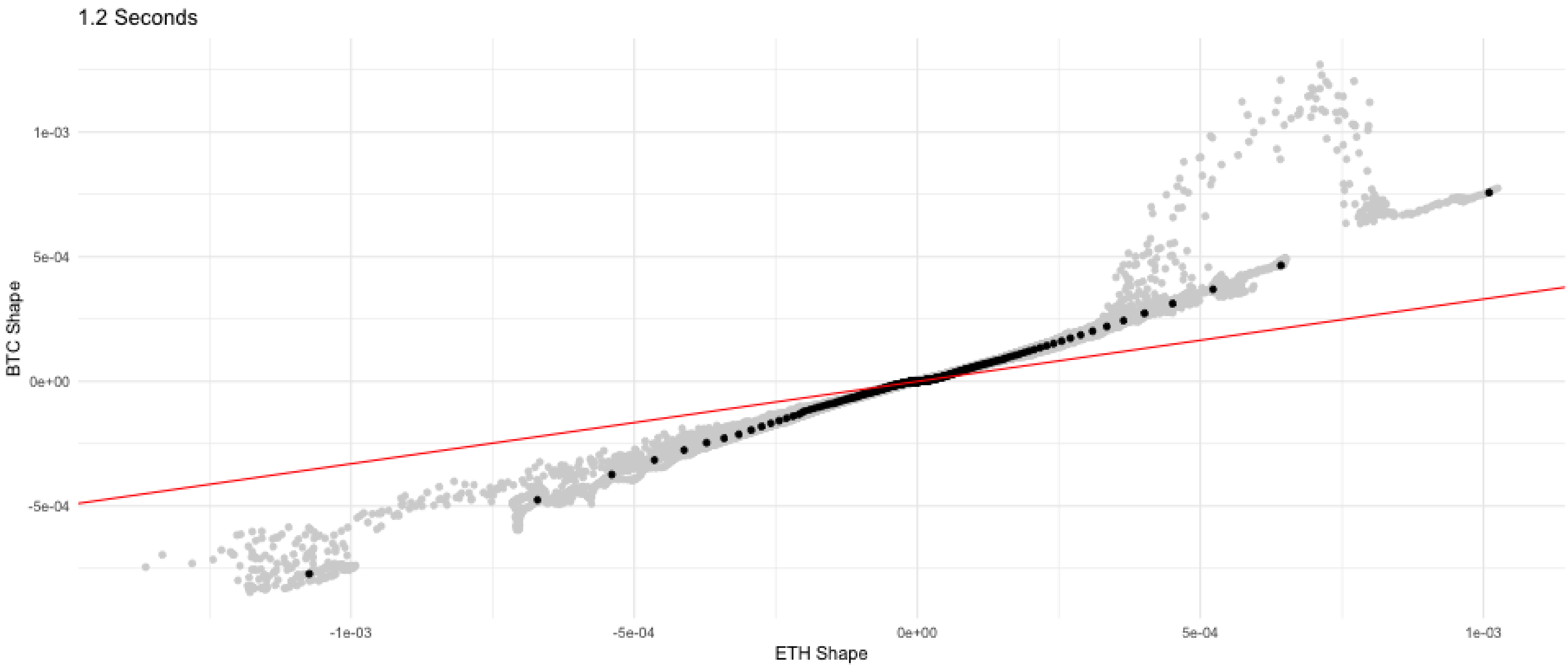
Based on the leptokutotic returns of Ether prices in Bitcoin terms, it is hard to know what to expect.

HF ETH Versus BTC QQ

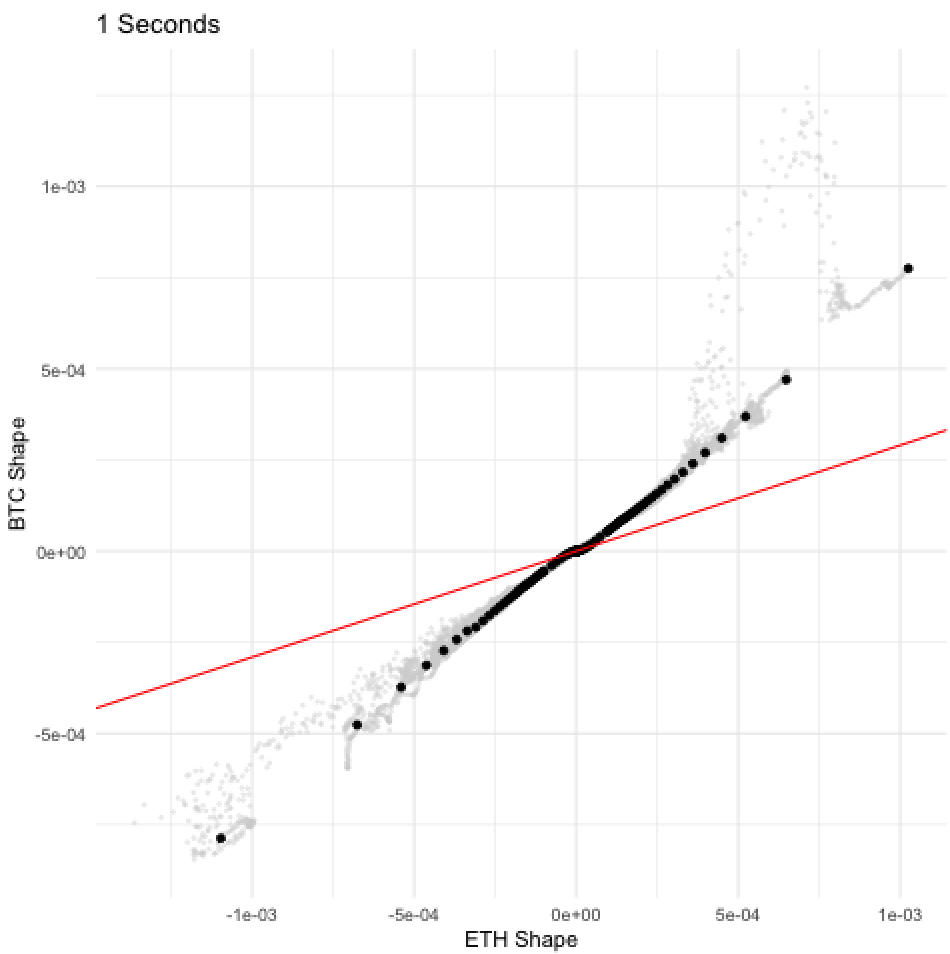
Time Scale:

5 200

1 21 41 61 81 101 121 161 200



Distribution of ETH Versus BTC (Cycling)



BTC is more fat-tailed than ETH

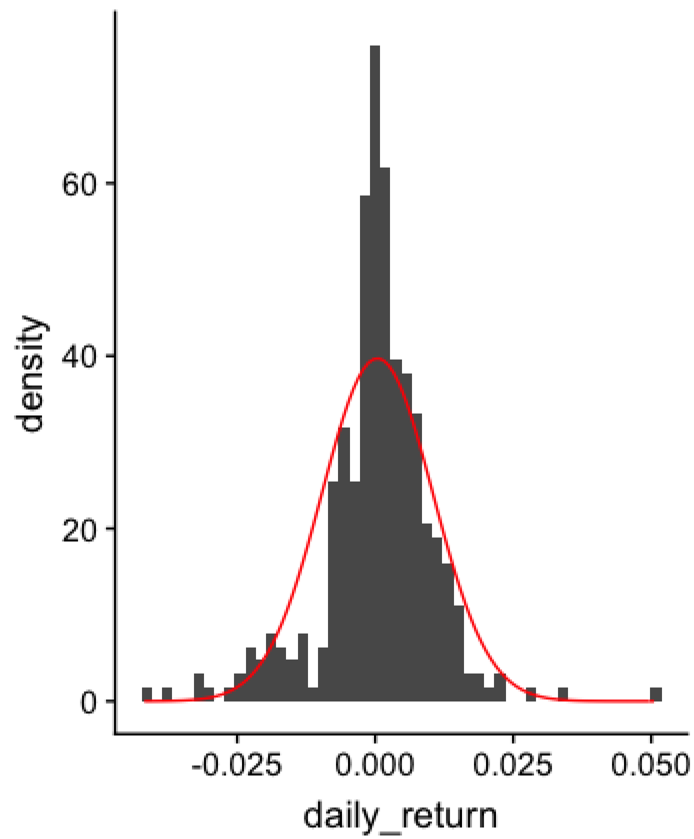
How Do These Compare With Other Assets?

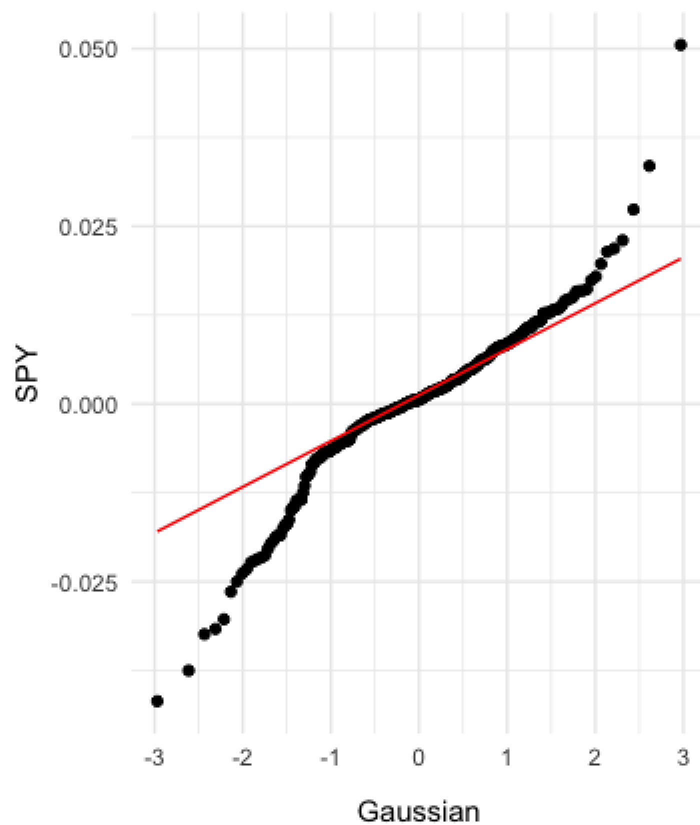
Comparisons With Other Assets Help Us Understand Implications

A natural question, since equities too are known to have fat-tailed returns, arises:

Are there qualitative differences between ETH and BTC returns and other asset returns?

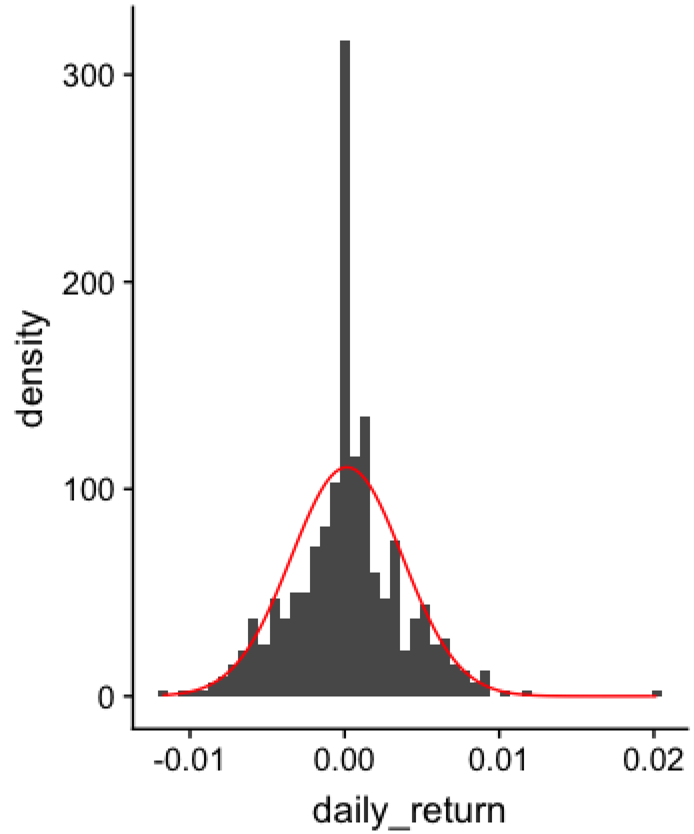
Broad Equity: SPY

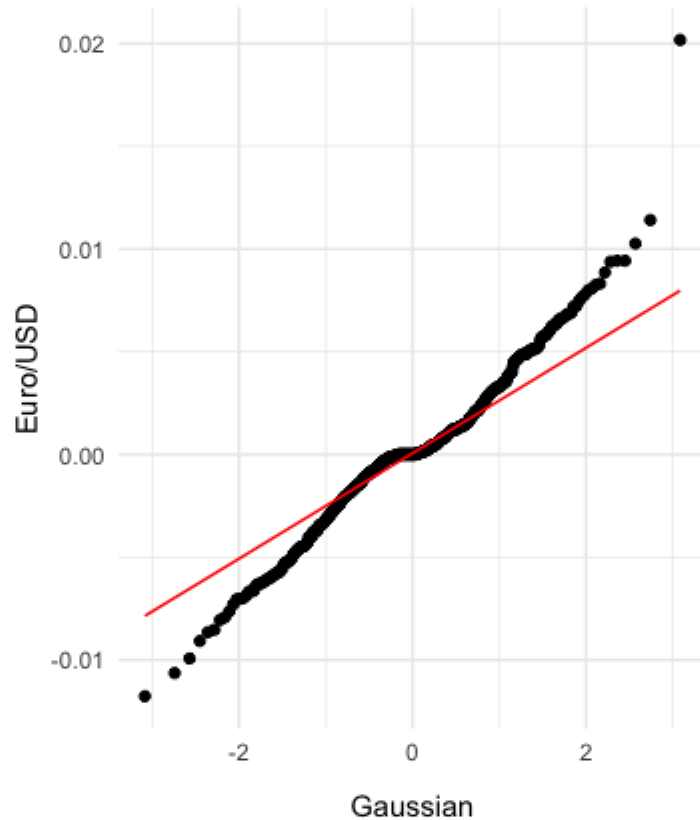




SPY returns are not gaussian

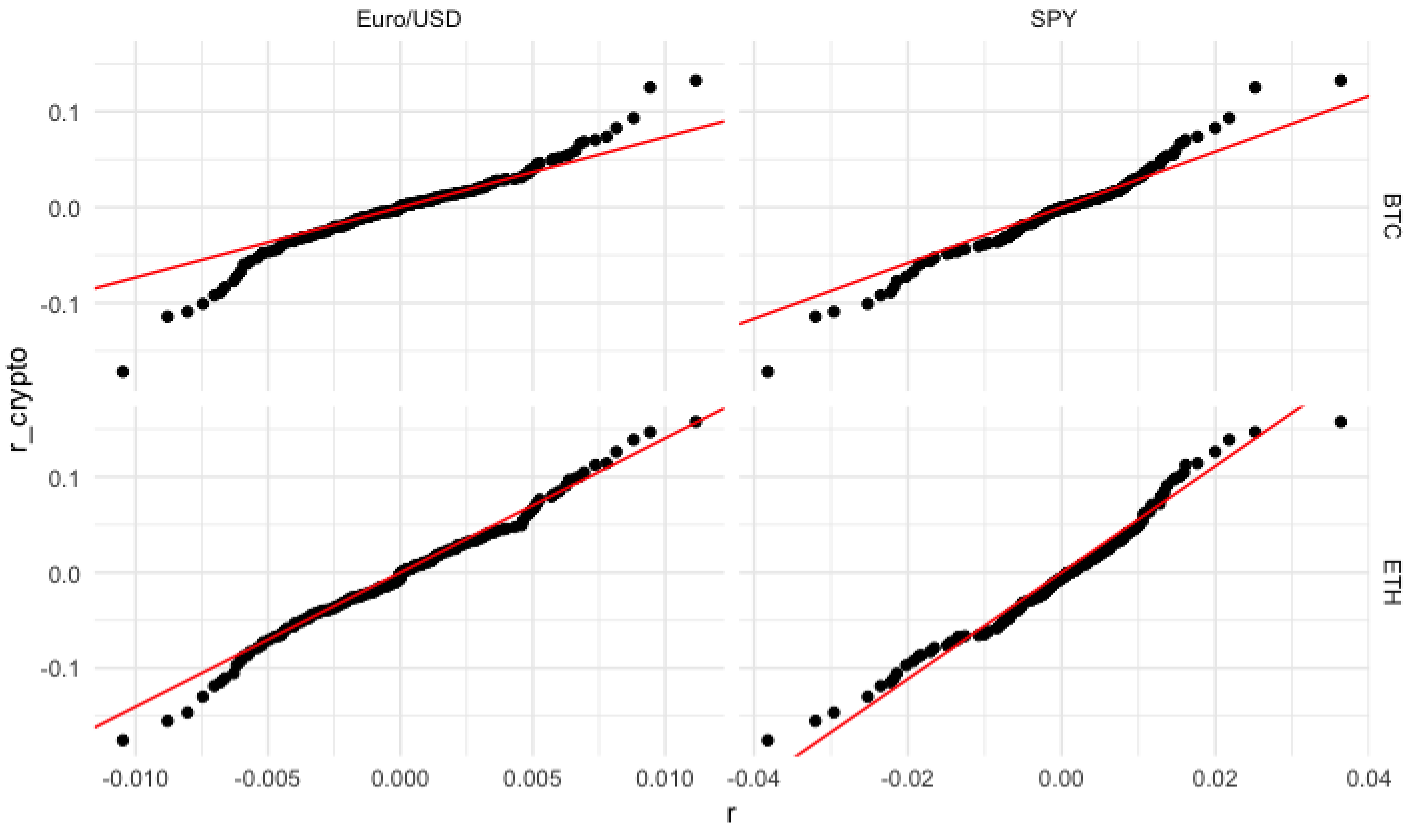
Liquid Currency: EUR





Euro currency returns are not gaussian

Differences To SPY And Euro



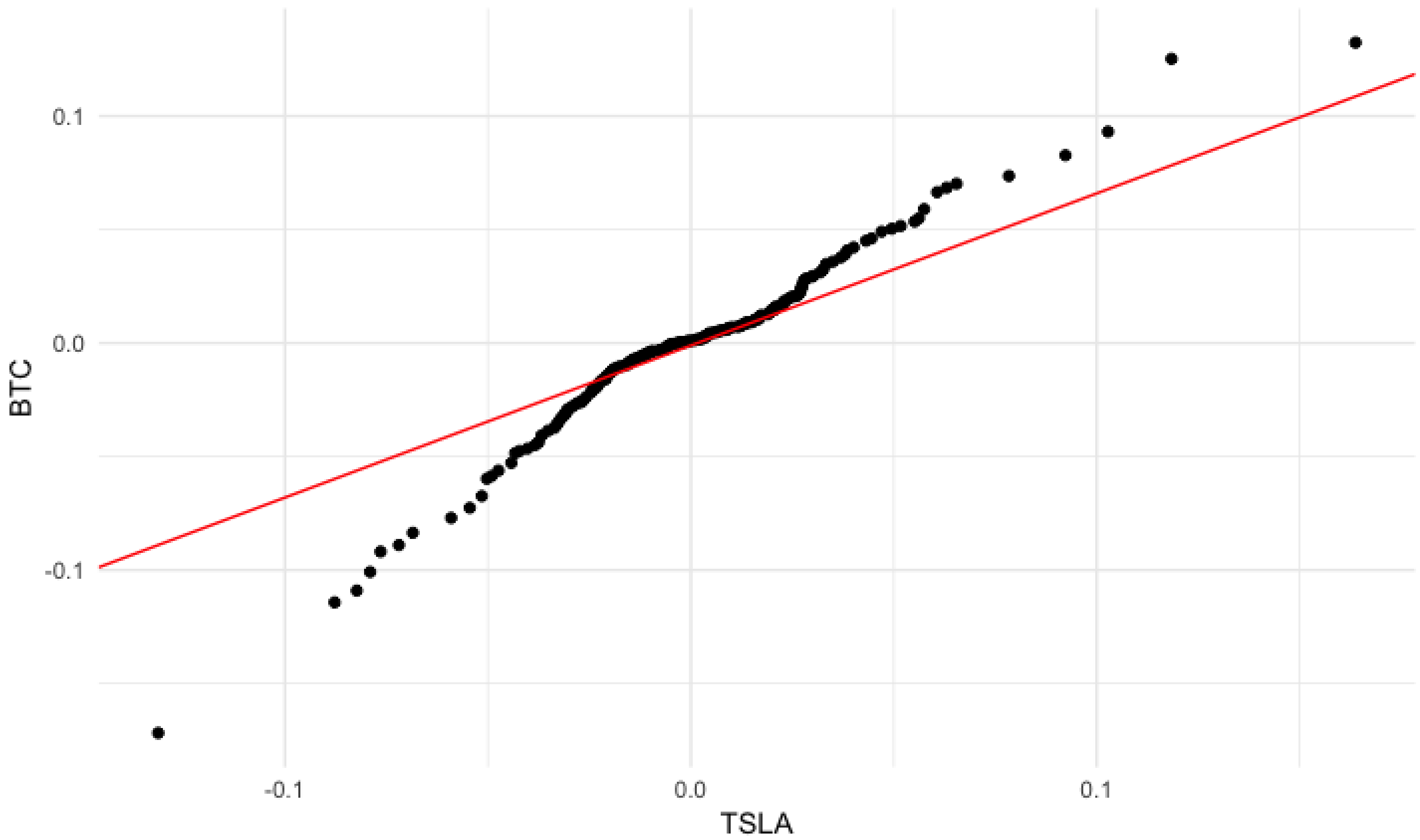
Daily ETH resembles SPY and Euro, BTC does not

How Do We Compare To Equities In General?

BTC returns are fatter-tailed than a broad market index. What about individual equities?

Bitcoin is popular in the public imagination, and experiences wild swings. What equity does that remind us of?

BTC Versus Tesla Equity Return, Daily



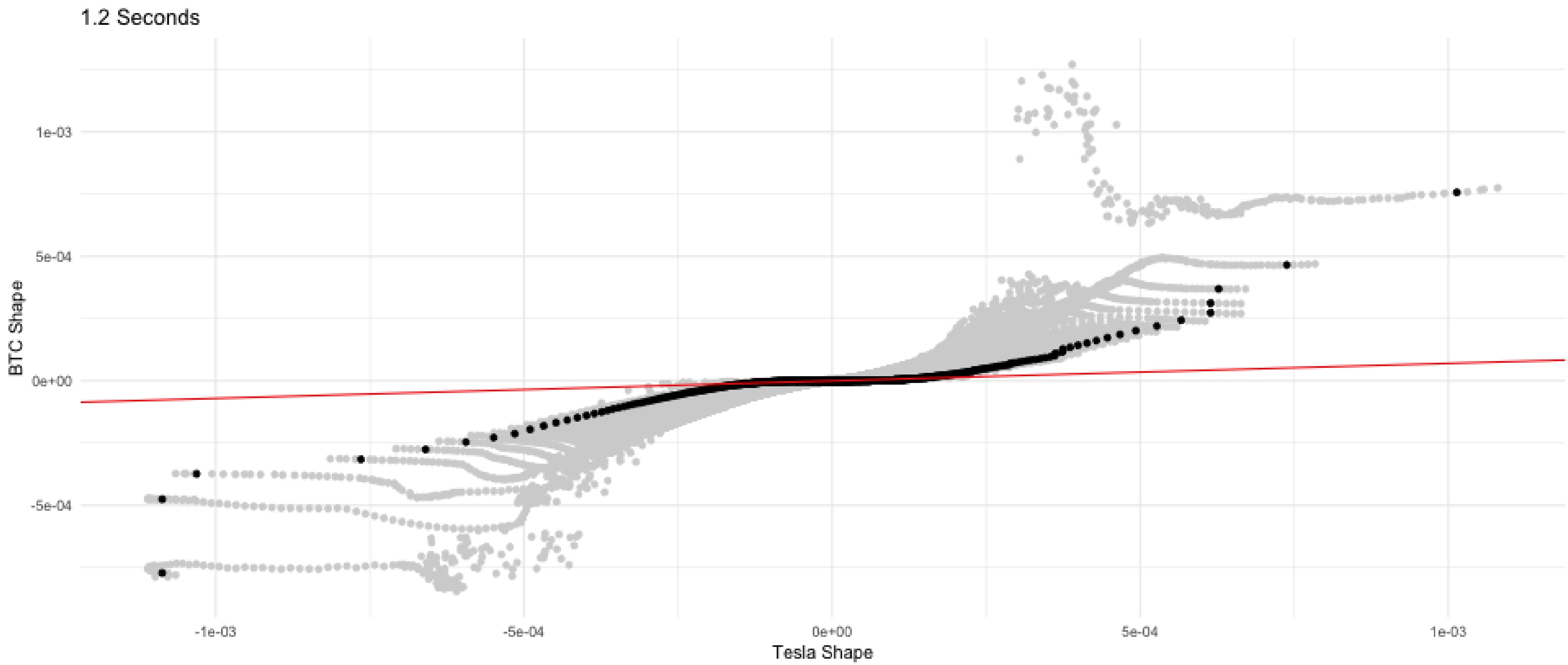
Bitcoin is more extreme than Tesla on daily data

HF BTC Versus TSLA QQ

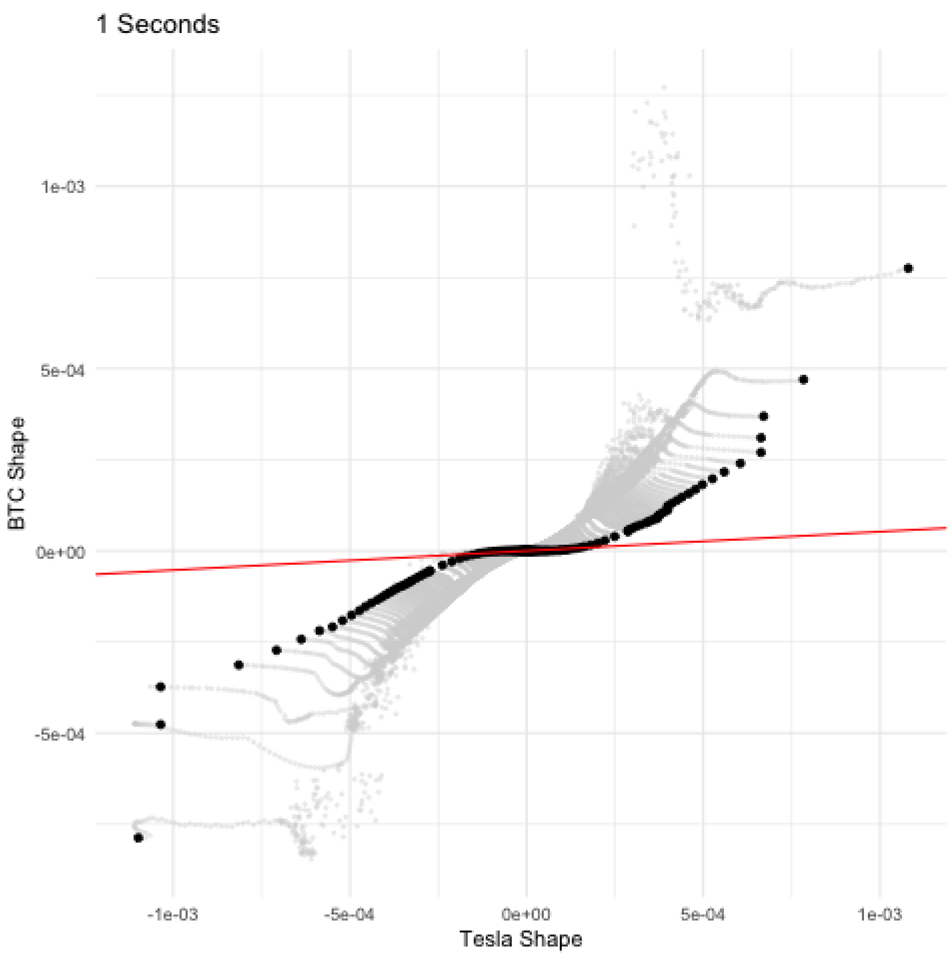
Time Scale:

5 200

1 21 41 61 81 101 121 161 200



Distribution of BTC Versus TSLA (Cycling)



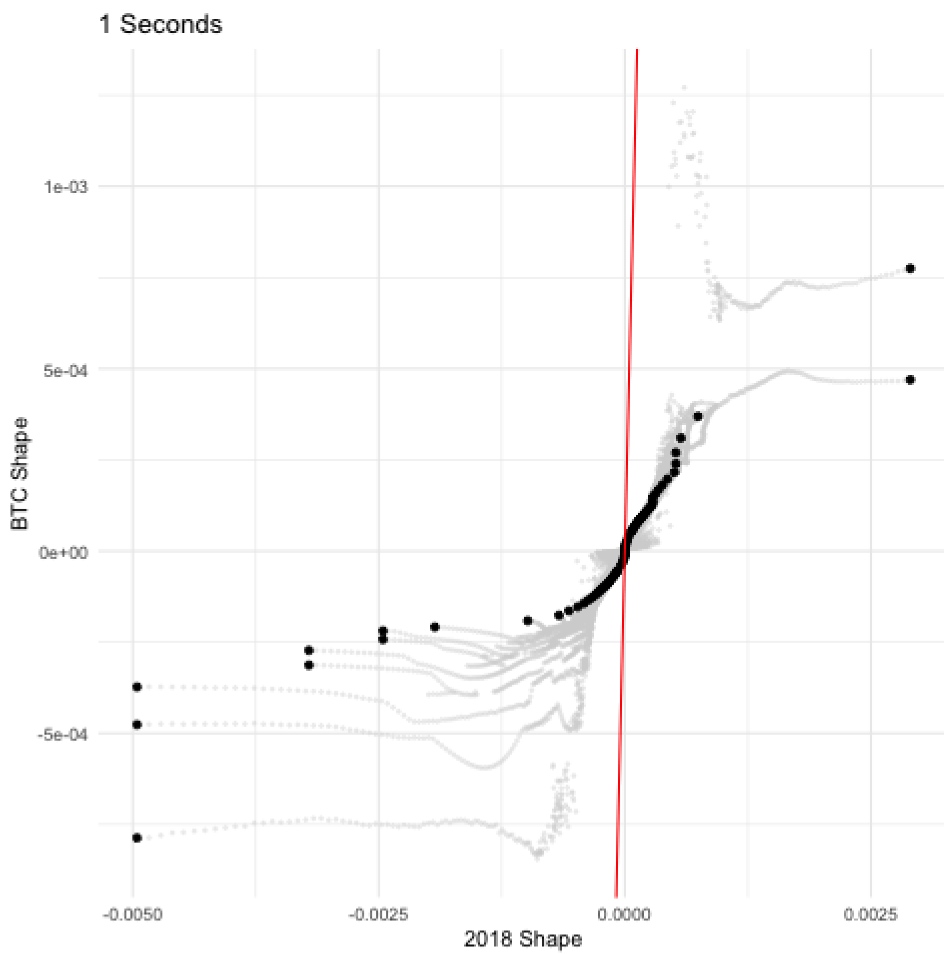
BTC is more fat-tailed than TSLA

Was 2018 Different?

In 2018, BTC was still very much in the public eye.

Were the return statistics any different at the time?

2018 Versus 2019



2018 Versus 2019

Market Making

Rather than investor returns, we might be thinking about trading possibilities.

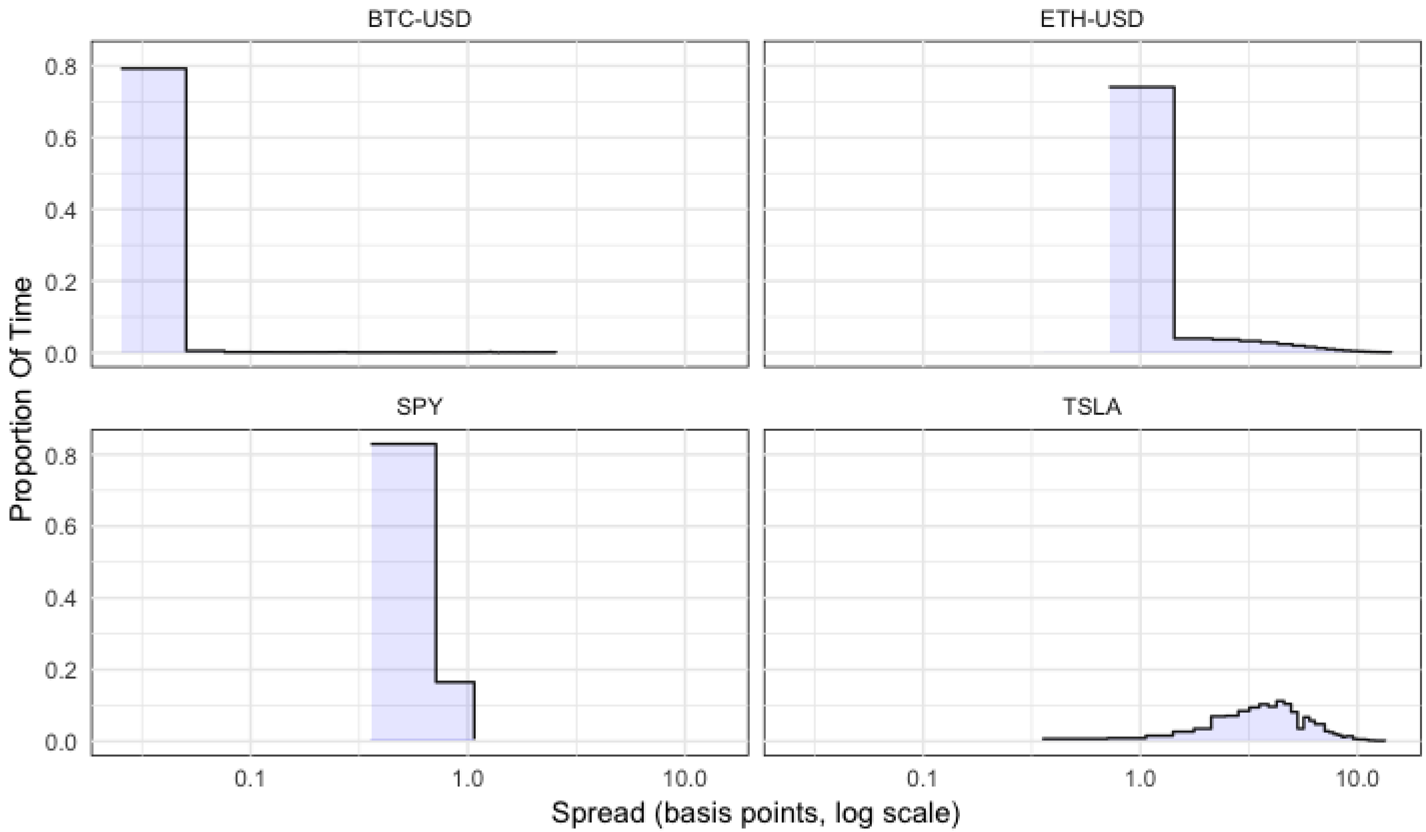
How much opportunity is there for high-frequency market-making?

Spread Distributions

A large portion of market-making profit is derived from the spread.

We can compare the distribution of spreads among our various assets.

Spread Distribution Plots



Crypto spreads are very narrow

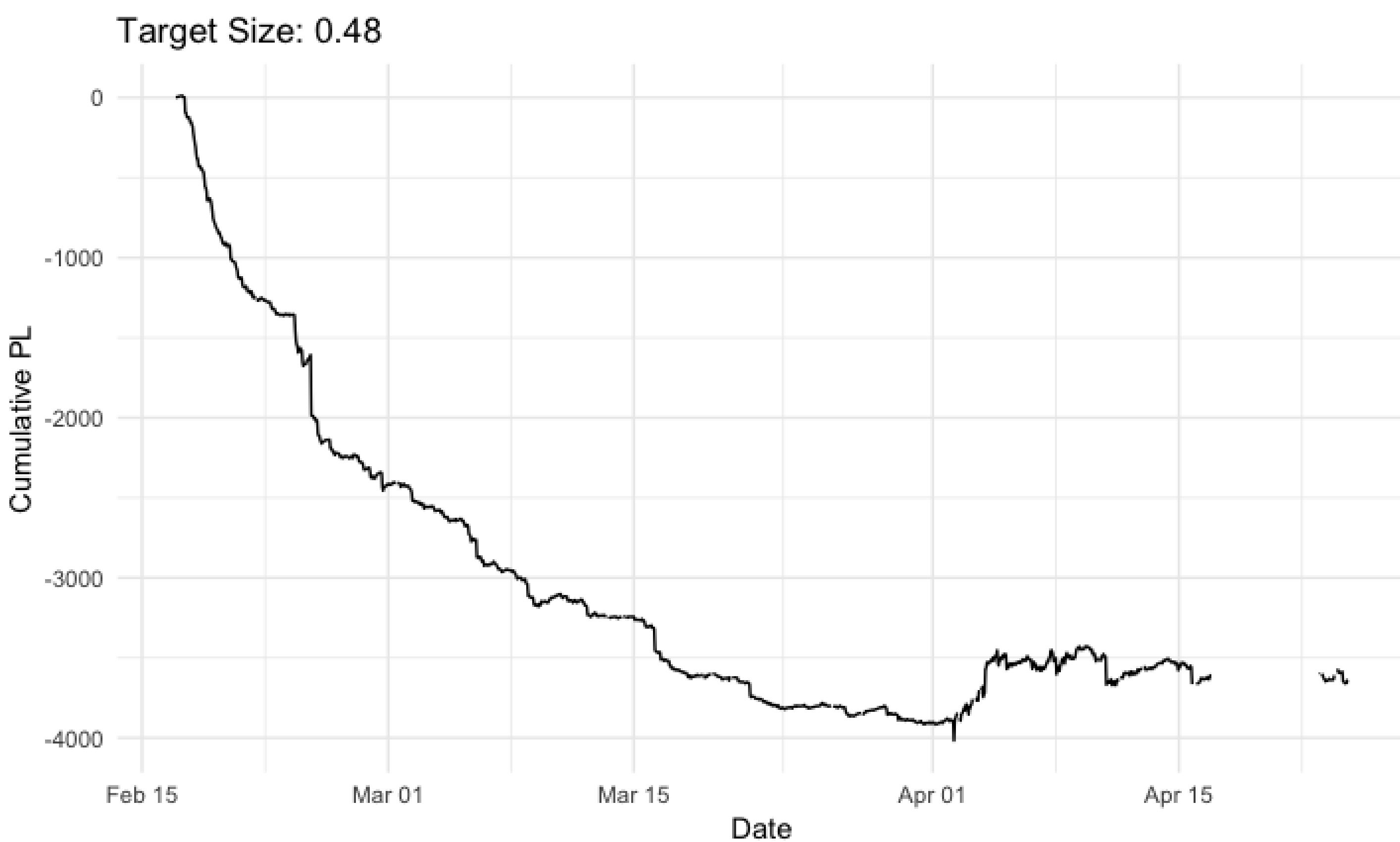
Simulated Participatory Market-Making

We can easily simulate a market maker who gets a portion of every trade, with constraints on trade size and position limits.

BTC Market Making Graph

BTC market making goes the wrong way

ETH Market Making Graph



ETH market making also goes the wrong way

Conclusions (Reprise)

- Cryptotoken returns (in USD terms) are fundamentally fatter-tailed than equity returns
- The fatter tails are *not* explained by (slow-moving) stochastic volatility
- Bitcoin returns are fatter-tailed than Ethereum returns
- This is true on both daily and high-frequency time scales
- Combined with tiny spreads, this means classic market-making is unprofitable