Supplementary Material

1 SUPPLEMENTARY METHODS

1.1 Text Preparation

1.1.1 Submissions filtered out

The following submissions were removed:

- Automated submissions authored by the following: 'AutoModerator', 'CommunityPoints', 'rBitcoinMod' and 'crypto_bot';
- Those consisting of duplicate text;
- Those containing just '[deleted]' or '[removed]'; and
- Blank submissions.

1.1.2 Text processing

All text was placed into lower case and strings of 50 or more word characters (too long to represent a word) were removed. The following details the approach to standardising synonymous words (e.g. 'BTC' and 'bitcoin') and replacing terms of multiple words ('smart contract') with single words ('smartcontract'). We accounted for words being separated by whitespace characters ('smart contract') and hyphens ('smart-contract'); and for spelling variants (decentralised and decentralized).

Currency codes

Cryptocurrency codes were replaced by the name of the associated cryptocurrency (Table S1). References to '1BTC' or '1 XBT' became '1 bitcoin'; 'ETH' and 'ether(s)' became 'ethereum'. This was applied to the top 10 cryptocurrencies by market capitalisation and/or liquidity (13:41 UTC; 21 May 2019): Bitcoin, Ethereum, Ripple, Bitcoin Cash, Litecoin, Binance Coin, Tether, Stellar, Cardano and Tron. The cryptocurrencies EOS, Matic Network and NEO did not have a distinct currency code. The abbreviation SAT was further replaced with satoshi (Bitcoin Wiki, 2018). Other cryptocurrencies were added to this list where highlighted by previous runs of the methodology: Golem, Verge, Ethereum Classic, Bitcoin Unlimited, Iconomi, Distributed Credit Chain, UChain, Bancor, Maker DAO, DIGIX and Auctus. Before this, references to US dollars were standardised.

Improvement proposals

The following improvement proposal references were standardised: 'bitcoin improvement proposal(s)' and 'bips' were converted to 'bip'; 'ethereum improvement proposal(s)' and 'eips' were converted to 'eip'; and 'ethereum request(s) for comment(s)' was changed to 'erc'. References to the same numbered proposal were standardised through removing the gap between the proposal type ('erc') and number of proposal ('20'). Hence, 'erc-20', 'erc 20' and 'erc20' all became 'erc20'.

Cryptocurrency, financial, regulator and nationality words

We standardised ethereum-related (Table S2); bitcoin-related (Table S3); cryptocurrency-related (Table S4); and finance-related (Table S5) terminology.

Table S1. Replacing currency code with associated name. Conversion of ETC into 'ethereumclassic' was conducted prior to lower-case conversion to prevent confusion with *et cetera* ('etc.').

Replacing Term	Terms Replaced
dollarmarkersymbol	'(us/u.s.) dollar(s)'; 'usd'; '$'
ethereum	'eth'; 'ether(s)'
ethereumclassic	'ETC'; 'ethereum classic'
bitcoincash	'bch'; 'bitcoin cash'; 'bcash'
bitcoin	'btc'; 'xbt'; 'bitcoins'
satoshi	'sat(s)'; 'satoshis'
tron	'trx'
ripple	'xrp'
stellar	'xlm'
cardano	'ada'
litecoin	'ltc', 'litecoins'
golem	'gnt'
tether	'usdt'
binancecoin	'bnb'; 'binance coin(s)'
verge	'xvg'; 'verge currency'
bitcoinunlimited	'bu'; 'btu'; 'bitcoin unlimited'
iconomi	'icn'
distributedcreditchain	'distributed credit chain'; 'dcc'
uchain	'ucn'
bancor	"bancor(')(s) network token"; 'bnt'
makerdao	'maker dao'; 'dai'
digix	'dgx'; 'dgd'; 'digix dao'; 'digix gold token(s)'
auctus	'auc'

Table S2. Ethereum concepts standardised.

New Term	Words Replaced
smartcontract	'smart contract(s)'
evm	'ethereum virtual machine'
dapp	'decentralized application(s)'; 'dapp(s)'; 'dap(s)'
dao	'decentralized autonomous organization(s)'; 'dao(s)'
dac	'decentralized autonomous corporation(s)'; 'dac(s)'
ico	'initial coin/token offering(s)'; 'token generation event(s)'; 'ico(s)'; 'ito(s)'; 'tge(s)'
eea	'enterprise ethereum alliance'

Table S3. Bitcoin concepts standardised. The term 'lightening' is a common mistake in spelling 'lightning' (Ticak, 2016).

New Term	Words Replaced
ln	'light(e)ning network(s)'
segwit	'segregated witness'; 'sw'
segwit2x	'b2x'; 's(w)2x'; 's(w)2mb'; 'segwit 2mb'; 'segwit2mb'; 'segwit 2x'
nya	'bitcoin scaling agreement at consensus 2017'; 'new york agreement'

Table S4. Cryptocurrency concepts standardised.

New Term	Words Replaced
cryptoasset	‘crypto currency/ies’; ‘crypto asset(s)’; ‘cryptocurrency/ies’
delegatedproofofstake	‘delegated proof of stake’; ‘dpos’
proofofstake	‘proof of stake’; ‘pos’
proofofwork	‘proof of work’; ‘pow’
proofofauthority	‘proof of authority’; ‘poa’
byzantinefaulttolerance	‘byzantine fault tolerance’; ‘bft’
directedacyclicgraph	‘directed acyclic graph(s)’; ‘dag’
storeofvalue	‘store of value’; ‘sov’
mediumofexchange	‘medium of exchange’; ‘moe’
unitofaccount	‘unit of account’; ‘uoa’
cpu	‘central processing unit(s)’; ‘cpus’
gpu	‘graphics processing unit(s)’; ‘gpus’
asic	‘application specific integrated circuit(s)’; ‘asics’
asicboost	‘asic boost’
uasf	‘user activated soft fork(s)’
hashrate	‘hash power’; ‘hash rate’
twofactorauthentication	‘two/2/multi factor authentication’; ‘2fa’
ddos	‘distributed denial of service’
ipfs	‘interplanetary file(s) system’
pki	‘public key infrastructure’
publickey	‘public key(s)’
privatekey	‘private key(s)’
nonce	‘number used only once’
hardfork	‘hard fork’; ‘hf’
softfork	‘soft fork’
hd	‘hierarchical deterministic’
explain	‘eli5’
fud	‘fear(,) uncertainty(,) (and) doubt’
ai	‘artificial intelligence’
transaction	‘tx’
txid	‘transaction id(entification)’
tpsec	‘transaction(s) per second’; ‘tps’

Table S5. Finance concepts and nationalities standardised.

New Term	Words Replaced
etf	‘exchange traded fund(s)’
etp	‘exchange traded product(s)’
otc	‘over the counter’
dex	‘decentralised exchange(s)’
cex	‘centralised exchange(s)’
pumpanddump	‘pump(s/ed/ing) and dump(s/ed/ing)’
marketcap	‘market cap(italisation)(s)’
larger	‘bigger’; ‘larger’
technicalanalysis	‘ta’; ‘technical analysis’
fundamentalanalysis	‘fa’; ‘fundamental analysis’
kyc	‘know your customer/client’
sec	‘securities and exchange commission’
ftc	‘federal trade commission’
cftc	‘commodity futures trading commission’
fdic	‘federal deposit insurance corporation’
doj	‘department of justice’
g20	‘group of twenty/20’
pboc	“people’s bank of china”; ‘pbc’
cboe	‘chicago board options exchange’
ice	‘intercontinental exchange’
p2p	‘peer to/2 peer’
korea	‘(south) korea(n)’
france	‘french’
china	‘chinese’

Text removed

The following were removed respectively: URLs; HTML tags (e.g. '&amp'); the new line character ('\n'); references to deleted text ('[removed]' and '[deleted]'); greetings ('hey', 'hi' and 'hello') and non-ASCII text (e.g. Cyrillic alphabet or emoticons). Punctuation and apostrophes were removed unless these were inside words to indicate abbreviations (e.g. 'o'clock').

Creating lists of words from strings of text

The processed text was tokenised into word lists using Python package NLTK version 3.3 and its associated download 'punkt'. 'Stopwords' were then removed using the list provided by NLTK, supplemented by abbreviations for 'not' ("n't"); 'I am' ('im', "i'm"); 'you are' ("you're", 'youre'); '(s)he is' ('(s)hes', "(s)he's"); 'they are' ('theyr', "they'r", "they're", 'theyre'); and 'we are' ('wer', "we'r", "we're"). Words were also removed that contained no letters, thus deleting any numbers, along with references to thousands ('5k' or '14k'), millions ('5m', '1m'), multiples ('10x'), ranks ('1st', '2nd' or '4th') and images ('img'). Words were lemmatised using NLTK's 'WordNetLemmatizer', and stemmed using 'SnowballStemmer'. The 'snowball' stemmer was selected in being least likely to treat words of the same concept differently or words of a different concept the same (Jivani, 2011). Table S6 lists lemmatised and stemmed words that were standardised as they referred to similar concepts.

Table S6. Lemmatised and stemmed words standardised.

New Word	Word Replaced
'mine'	'miner'
'newbi'	'noob'; 'n00b'; 'newb'
'buy'	'purchas'
'ad'	'advertis'; 'advert'
'mew'	'myetherwallet'; 'myetherwalletcom'; 'wwwmyetherwalletcom'
'verif'	'verifi'; 'verif'
'repli'	'respons'
'might'	'mayb'
'partner'	'partnership'

1.2 Further details on finding the context of concepts

The context of a concept was examined through finding the top five most common words occurring in submissions containing at least one word from that concept. We removed the following words from the text before running the analysis as these did not aid in the interpretation of the concept: the name of the cryptocurrency being analysed, 'account', 'actual', 'add', 'address', 'ago', 'alreadi', 'also', 'amount', 'anyon', 'appli', 'back', 'blockchain', 'come', 'communiti', 'could', 'crypto', 'cryptoasset', 'current', 'day', 'differ', 'drive', 'end', 'even', 'everi', 'exchang', 'extra', 'feel', 'find', 'first', 'get', 'give', 'go', 'group', 'happen', 'howev', 'includ', 'keep', 'know', 'let', 'like', 'look', 'lot', 'make', 'mani', 'may', 'money', 'much', 'multipl', 'need', 'next', 'one', 'peopl', 'pleas', 'put', 'rememb', 'right', 'run', 'say', 'see', 'similar', 'someth', 'start', 'still', 'take', 'talk', 'thing', 'think', 'time', 'two', 'use', 'user', 'want', 'way', 'whole', 'work', 'would', 'year' and 'yet'. If two or more words were in the same percentage of submissions (rounded to two decimal places), we treated such words as being ranked equally.

2 SUPPLEMENTARY RESULTS

2.1 Local extrema

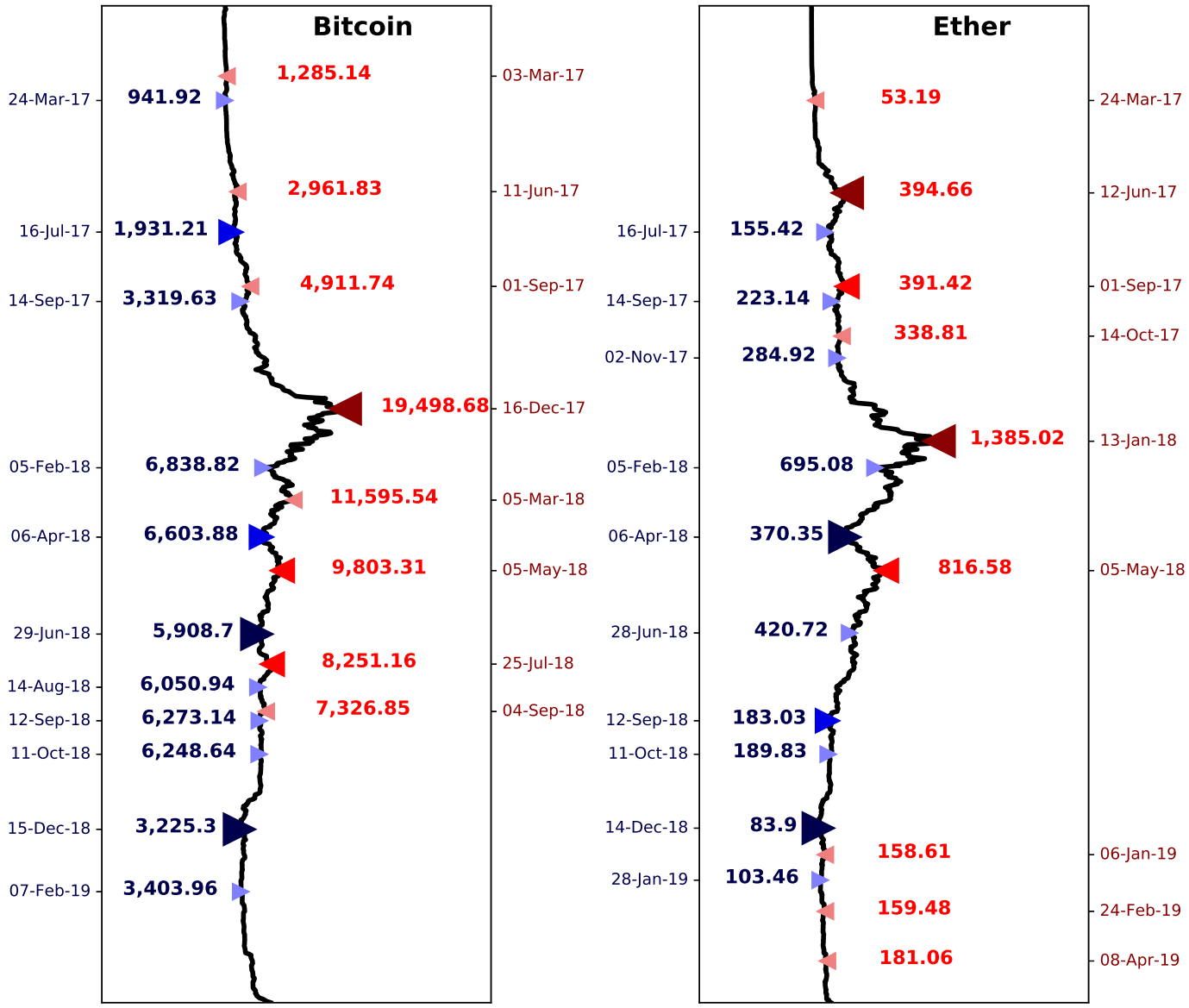

Figure S1. Comparison of ether and bitcoin US Dollar Price Local Extrema (1 January 2017 to 14 May 2019). Local minima indicated by blue '▷' and local maxima by red '◁'. Smallest, lightest-coloured symbols indicate most extreme (highest or lowest) price for 28 days (4 weeks or about 1 month) before and after; next size 56 days (8 weeks or about 2 months); and largest, darkest-coloured 84 days (12 weeks or about 3 months) before and after. Dates of minima on left and maxima on right.

2.2 Descriptive statistics

Table S7 shows descriptive statistics for the different price phases for Bitcoin and Ethereum.

Table S7. For each phase in the cryptocurrency price series, the number of days and submissions.

Bitcoin		
Phase	Days	Submissions
Rise	406	204344
Fall	168	86290
1	349	180898
2	51	48048
3	28	13290
4	32	12302
5	29	10156
6	55	17213
7	139	38700
8	30	8727
Ether		
Phase	Days	Submissions
Rise	242	61010
Fall	306	68034
1	162	30328
2	164	54372
3	51	24037
4	83	27552
5	29	6645
6	223	40482

2.3 Multi-phase concepts

The percentage changes in frequency across prices phases for each multi-phase concept are listed in Table S8 for Bitcoin and Table S9 for Ethereum.

Table S8. For each Bitcoin multi-phase concept, the percentage change in frequency with each shift to falling prices. Frequency was measured as proportion of submissions containing that concept. Both concepts identified were associated with falling prices.

Concept	Phase 1 to 2	Phase 3 to 4	Phase 5 to 6	Phase 7 to 8
sale	42.55	25.09	15.44	32.02
market	12.09	22.81	10.73	30.29

Table S9. For each Ethereum multi-phase concept, the percentage change in frequency with each shift to rising (upper section) or falling (lower section) prices. The first three concepts were associated with rising prices, and the next seven concepts were associated with falling prices. Frequency was measured as proportion of submissions containing at least one word from that concept.

Concept	Phase 2 to 3	Phase 3 to 4	Phase 4 to 5	Phase 5 to 6
tax	61.67		21.95	
hit	37.91		11.63	
dollarmarkersymbol	12.53		0.36	
makerdao, stablecoin		209.09		4.48
bear, bearish, bull		75.22		35.49
ceo, cofound		63.73		25.57
market		26.21		32.04
project, team		24.52		23.01
game		23.61		23.11
featur		21.45		35.64

REFERENCES

Bitcoin Wiki (2018). Satoshi (unit). *Bitcoin Wiki* `https://en.bitcoin.it/wiki/Satoshi_(unit)`

Jivani, A. G. (2011). A Comparative Study of Stemming Algorithms. *International Journal of Computer Technology and Applications* 2, 1930–1938

Ticak, M. (2016). Lightening vs. Lightning—What's the Difference? *Grammarly* `https://www.grammarly.com/blog/lightening-vs-lightning/`