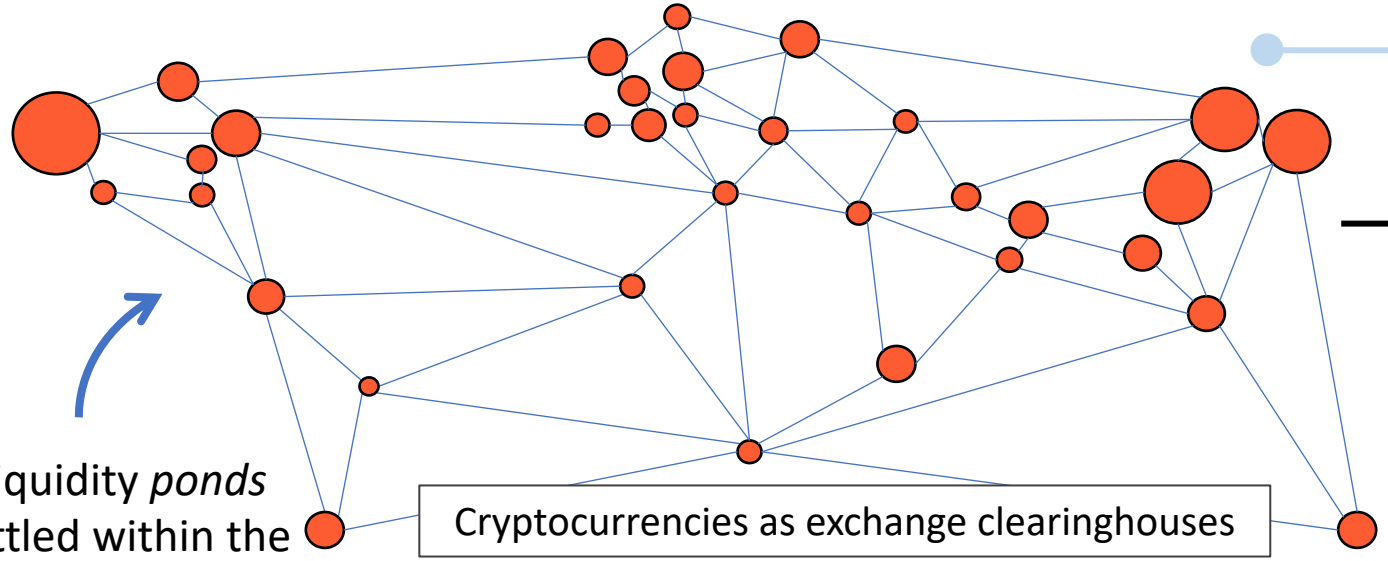
The Case for Better Data in Crypto

Chicago Universities Blockchain Summit

Nic Carter

CASTLE ISLAND VENTURES

COINMETRICS



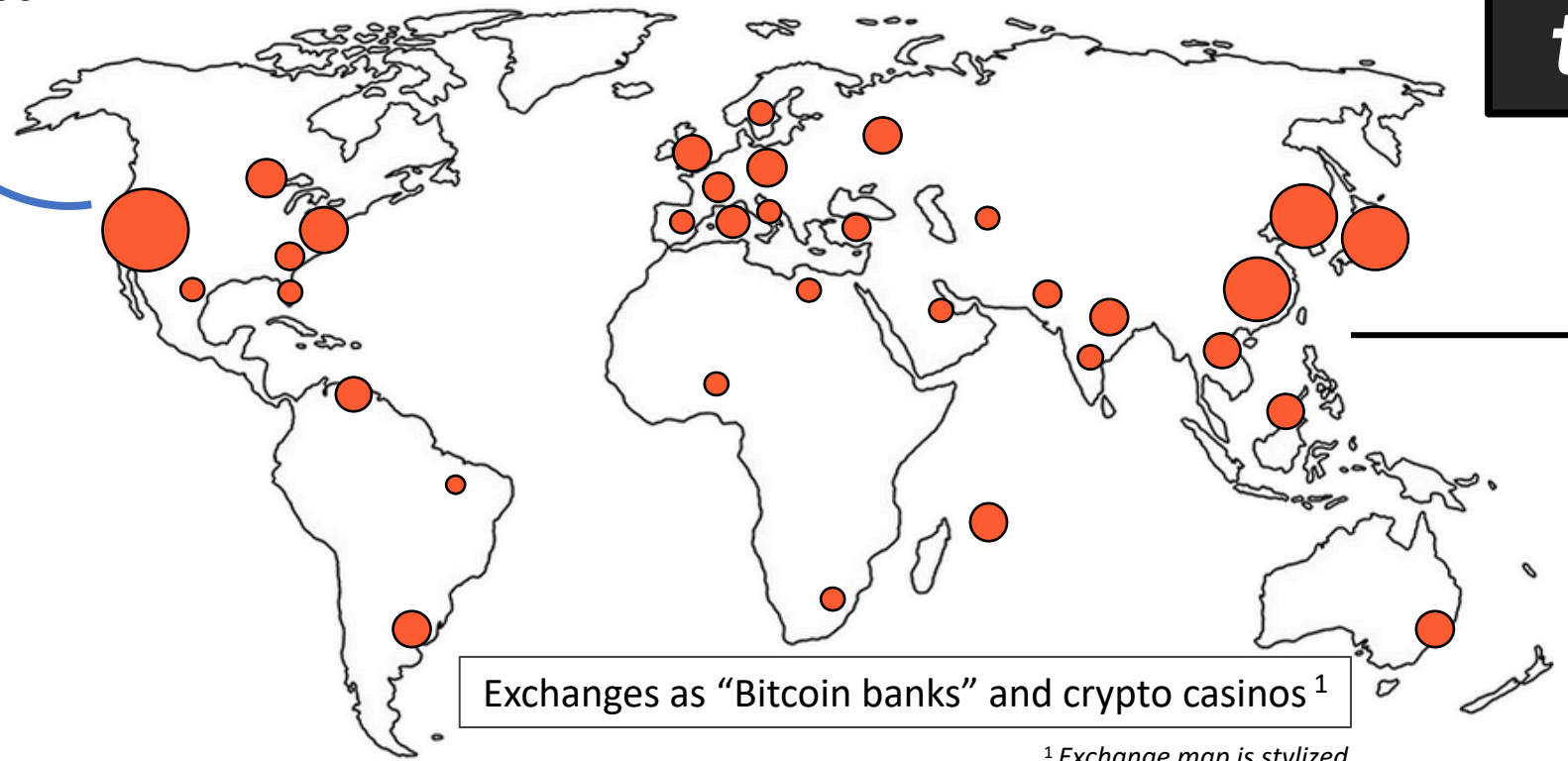
Cryptocurrencies as exchange clearinghouses

Liquidity *ponds* settled within the liquidity *pool*

Denoise publicly-available on-chain data

Understand the market

Denoise self-reported exchange data

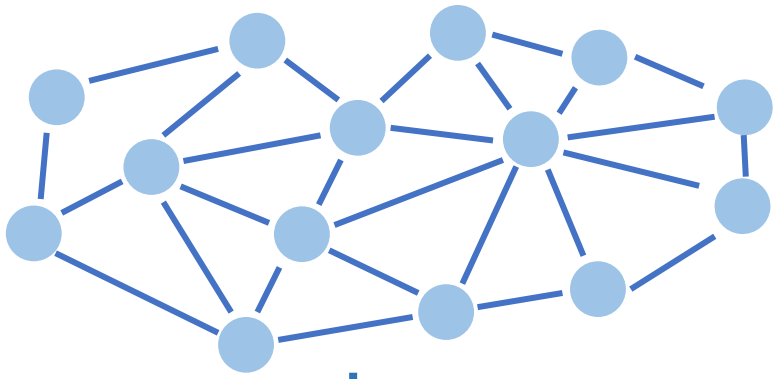


Exchanges as “Bitcoin banks” and crypto casinos [1]

[1] *Exchange map is stylized.*

The monetary stack

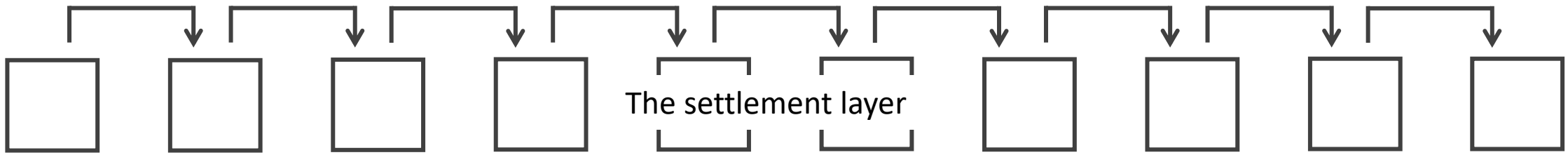
Limited data	Limited data	Virtually no data	Rich but nonstandard data
Overlay networks like Lightning	Sidechains like Liquid	Cash transactions	L2 transactions at exchanges



Liquid by Blockstream

coinbase
BINANCE
xapo
BitMEX
BITFINEX
LocalBitcoins.com

Periodic registry, settlement, and clearing at the base layer



Rich data but hard to parse and ambiguous

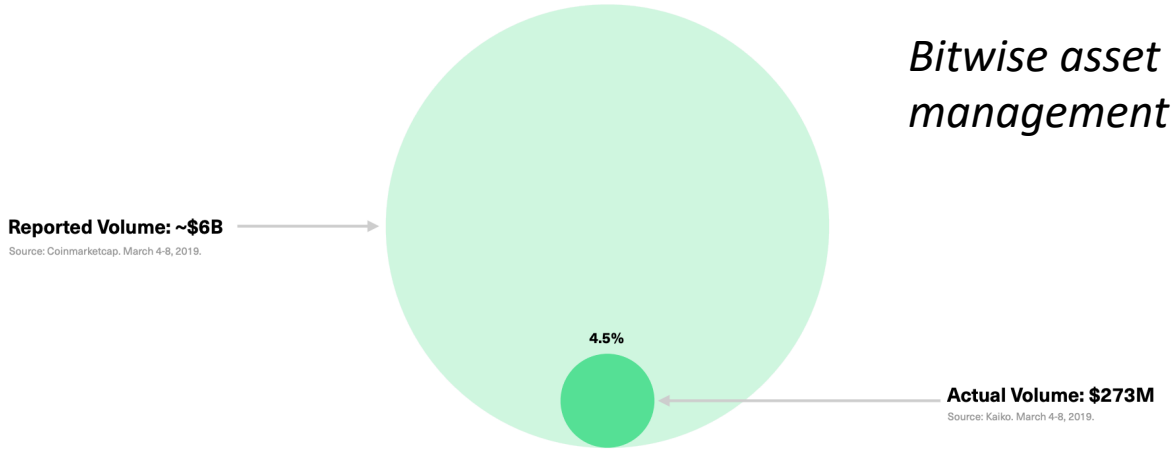
The sorry state of exchange data

Sylvain Ribes
Mar 10, 2018 · 11 min read

Chasing fake volume: a crypto-plague

In this piece I will expose why I believe more than $3 billion of all cryptoassets' volume to…

Total Volume Is Considerably Less Than Reported



Bitwise asset management

Cryptocurrencies

Report Says Most Crypto Trading Volume Is Suspicious

By Olga Kharif
March 19, 2019, 10:08 AM EDT

Binance Suffers From Wash Trading? And For $850, So Can You.

If the good exchange citizens can't escape the scourge of wash trading, is the industry in trouble?

By **Paul de Havilland** - May 6, 2019

Why do they do it?

Bitcoin Markets

Pair: All | Category: All | Fee Type: All | USD

#	Source	Pair	Volume (24h)	Price	Volume (%)	Category	Fee Type	Updated
1	BitMEX	XBT/USD	** $5,650,223,958	* $7,163.00	14.22%	Derivatives	No Fees	Recently
2	Negocie Coins	BTC/BRL	$1,989,586,804	* $7,[illegible].31	5.01%	Spot	Percentage	Recently
3	CoinBene	BTC/USDT	$1,048,582,883	$7,181.65	2.64%	Spot	Percentage	Recently
4	EXX	BTC/USDT	** $1,004,303,951	$7,241.20	2.53%	Spot	Transaction Mining	Recently
5	BW.com	BTC/USDT	$893,923,541	$7,186.42	2.25%	Spot	Percentage	Recently
6	OEX	BTC/USDT	$854,593,051	$7,199.76	2.15%	Spot	Percentage	Recently
7	Coineal	LTC/BTC	$805,022,731	$7,212.34	2.03%	Spot	Percentage	Recently
8	OKEx	BTC/USDT	$659,314,227	$7,185.12	1.66%	Spot	Percentage	Recently
9	Binance	BTC/USDT	$623,399,251	$7,178.96	1.57%	Spot	Percentage	Recently
10	Coinall	BTC/USDT	** $614,438,024	$7,182.61	1.55%	Spot	Transaction Mining	Recently
11	IDAX	BTC/USDT	$572,393,185	$7,187.20	1.44%	Spot	Percentage	Recently
12	ZBG	BTC/USDT	** $564,765,509	$7,184.82	1.42%	Spot	Transaction Mining	Recently
13	DigiFinex	BTC/USDT	$526,962,755	$7,191.20	1.33%	Spot	Percentage	Recently
14	Huobi Global	BTC/USDT	$504,641,439	$7,179.21	1.27%	Spot	Percentage	Recently
15	BitForex	BTC/USDT	$492,912,678	$7,183.01	1.24%	Spot	Percentage	Recently
16	DOBI Exchange	BTC/USDT	$454,302,085	$7,182.25	1.14%	Spot	Percentage	Recently
17	Coineal	BTC/USDT	$432,261,063	$7,180.56	1.09%	Spot	Percentage	Recently
18	FCoin	BTC/USDT	** $413,960,130	$7,181.61	1.04%	Spot	Transaction Mining	Recently
19	Bit-Z	BTC/USDT	$399,734,623	$7,184.80	1.01%	Spot	Percentage	Recently
20	DOBI Exchange	ETH/BTC	$386,988,867	$7,213.72	0.97%	Spot	Percentage	Recently
21	Coinall	LTC/BTC	** $384,781,114	$7,219.47	0.97%	Spot	Transaction Mining	Recently

Because this is the most lucrative advertising space in the (crypto) world

Exchange data is *Junk by Default*

- As long as incentives to spam are nonzero, spam will exist
- As long as gatekeepers don’t impose standards, exchanges will treat datafeeds like advertisements
- Exchanges must be assessed on a case by case basis – aggregates are guaranteed to include junk data
- Whitelist, don’t blacklist

- Why is crypto different?
  - Unregulated exchanges can spring up without approval or regulatory status, thanks to permissionless settlement rails

On-chain data is the ground truth

- Using on-chain data is the equivalent of predicting retail sales from satellite images of parking lots

- ***But*** – it's noisy, full of spam, hard to parse, and lacks metadata

Blockchains are an accounting revolution

Bitcoin full node

- Costs $300, less if you self-assemble
- Plug and play, no experience required to use
- Runs constantly, no operation required
- Proves validity of inbound transactions, integrity of bitcoin held, and audits the global supply of bitcoin

Gold full node

- Costs >$5000
- Requires specialized experience to operate
- Slow and unwieldy to use
- Proves integrity of small quantities of gold, does not prove anything about the global stock

Fiat full node

?

- Just trust us
- Just trust us
- Just trust us
- Just trust us

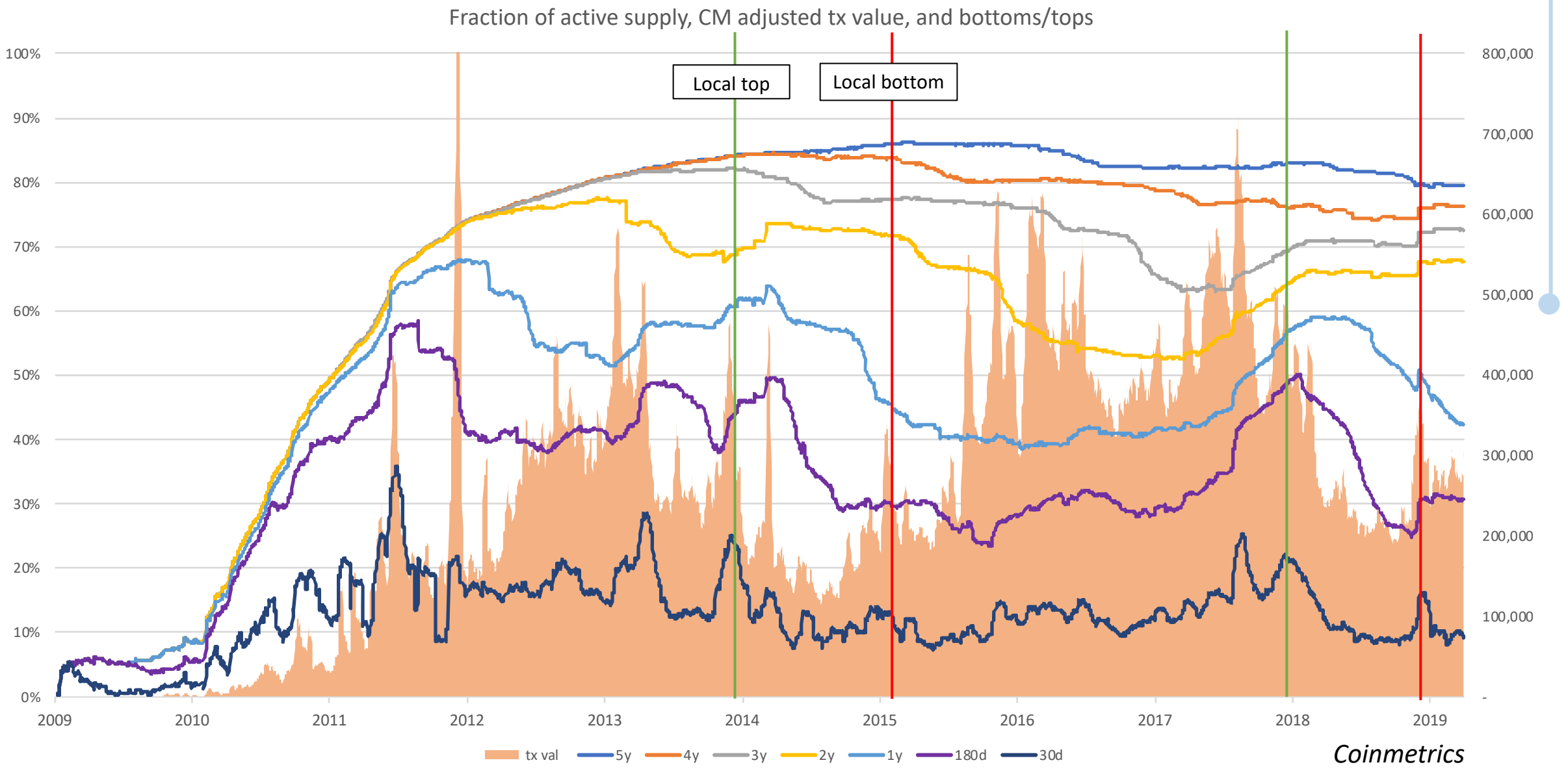
But someone has to interpret the data…

So how can you use on-chain data?

- Market timing / assessing our stage in the cycle
- Determining relative vibrancy and uptake of cryptocurrencies
- Ensuring that the chain has integrity and is secure
- Keeping issuers honest
- Auditing businesses that have on-chain components
- Evaluating the impact of upgrades, hard forks, or marketing initiatives
- In the future: granular financial reporting, continuous audits

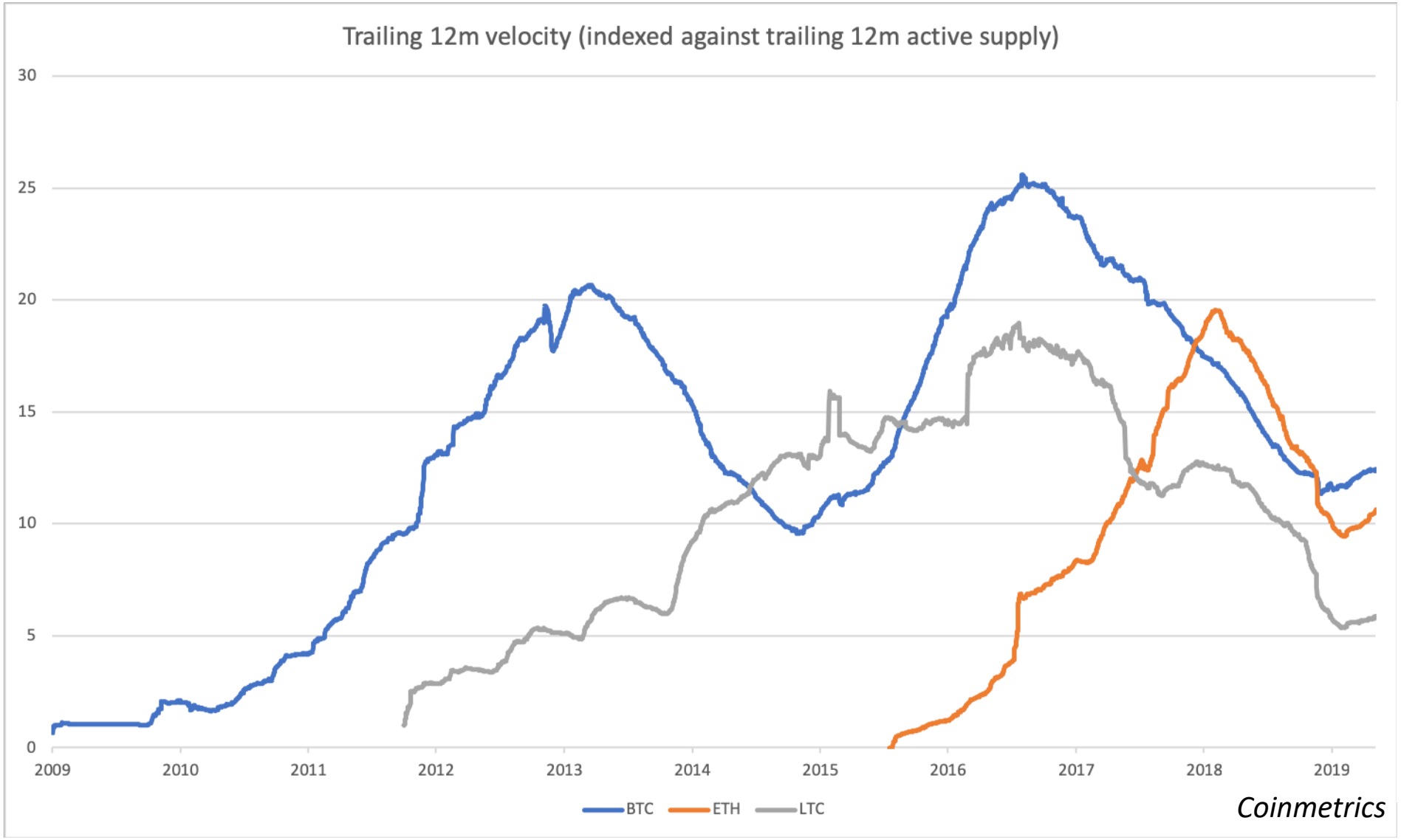
If you're not consulting the chain, you're doing it wrong!

Market cycles with supply cohorts and txn value



Coinmetrics

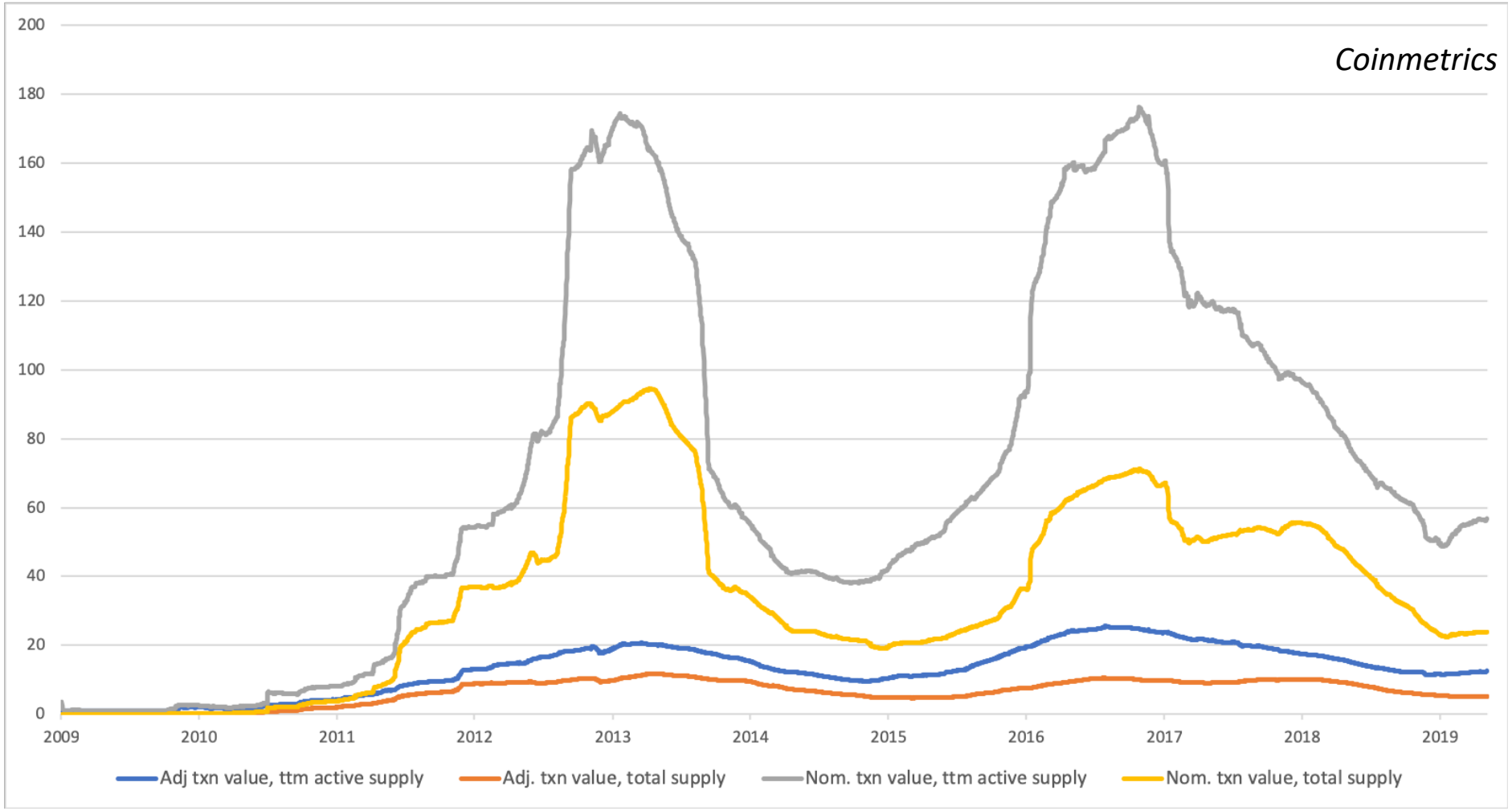
The Blockchain Business Cycle



- Bitcoin has been through two major business cycles, as usage of the network has waxed and waned
- Ethereum has been through just one, although it has synced up with Bitcoin recently
- Turnarounds in ttm velocity are generally very strong bottom signals

Formula: Sum of trailing 365d adjusted txn value / active supply (over the last 365d)

What's Bitcoin's real velocity?



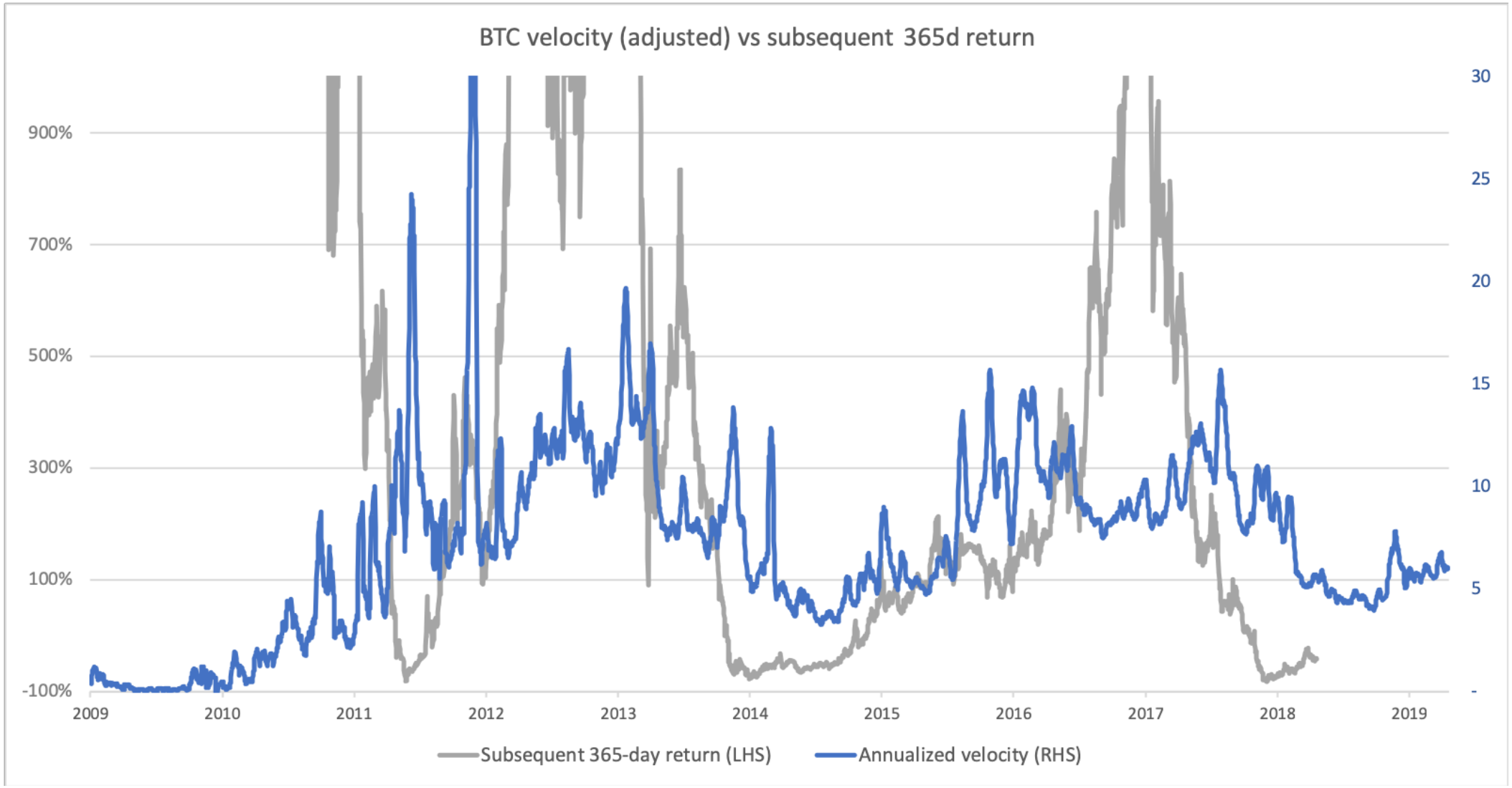
Bitcoin's annual velocity is between 5 and 50, depending on how you count it

I favor the most conservative: adjusted txn value and total supply

Velocity = transactional output / supply of units

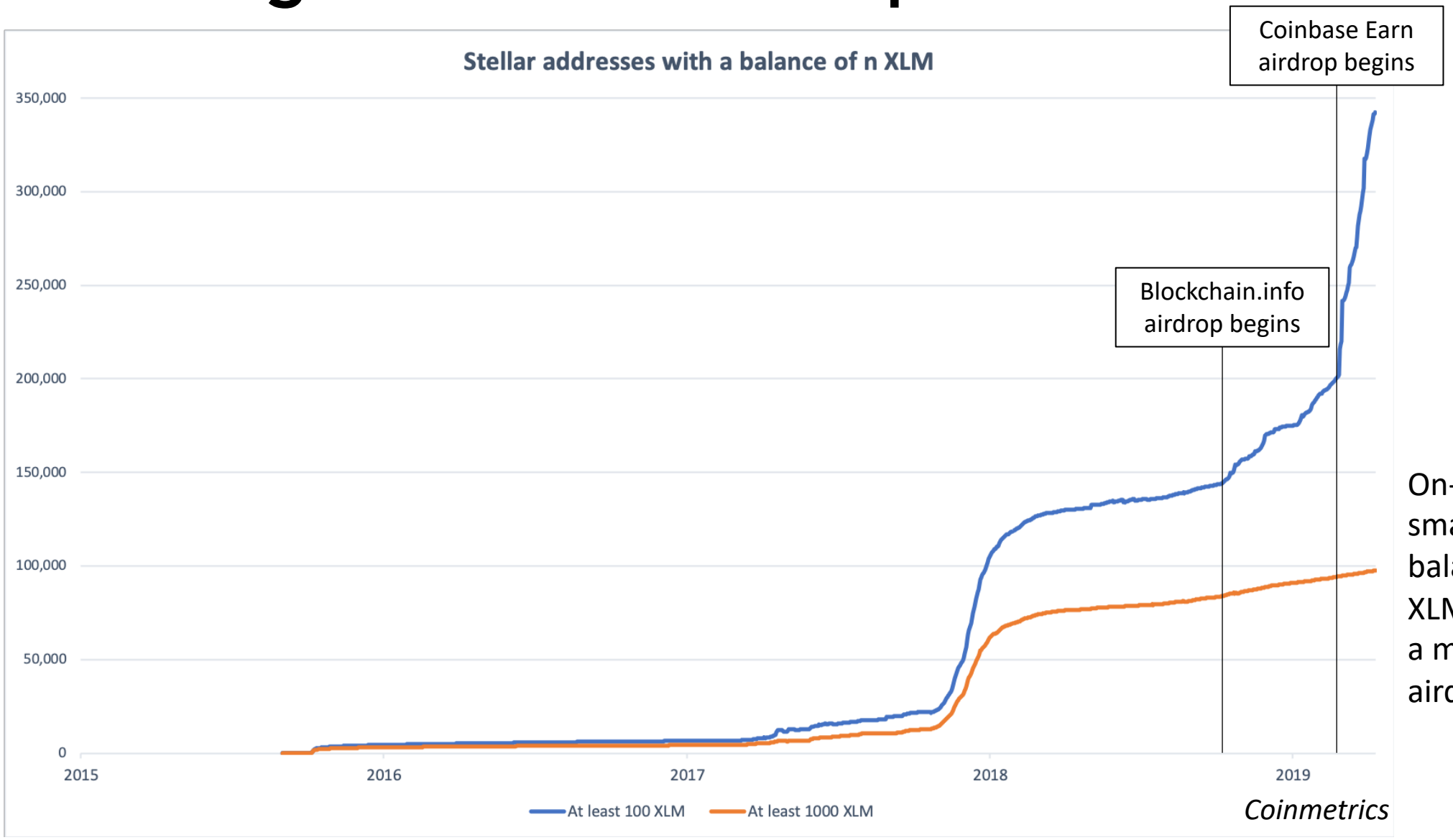
Which measure of output? Which measure of supply?

Does velocity analysis work?



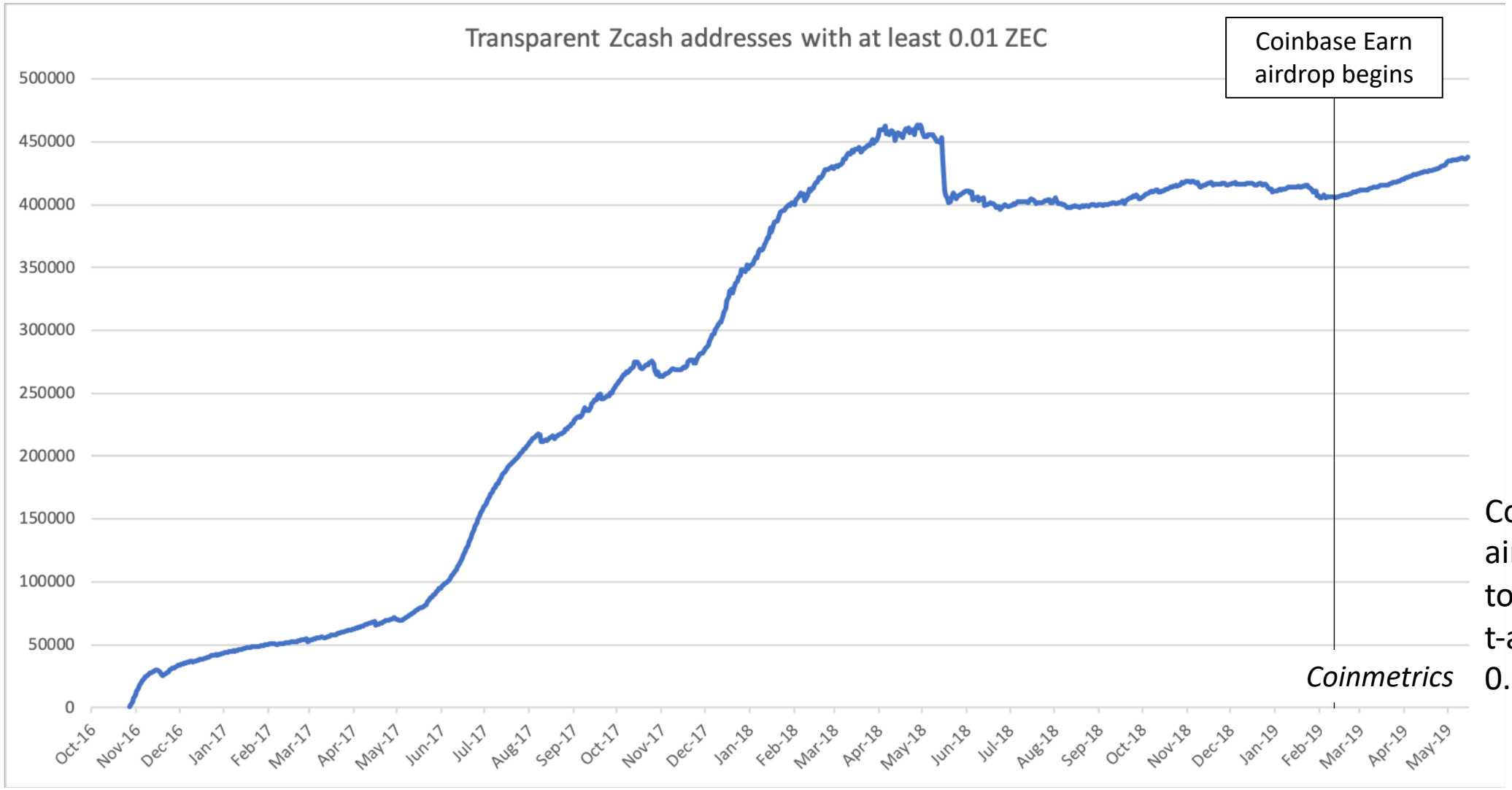
The jury is still out on this one…

Evaluating the Stellar Airdrops



On-chain wallets with small but meaningful balances (at least 100 XLM or ~$12) showed a marked increase as airdrops began

Evaluating the Zcash Airdrop

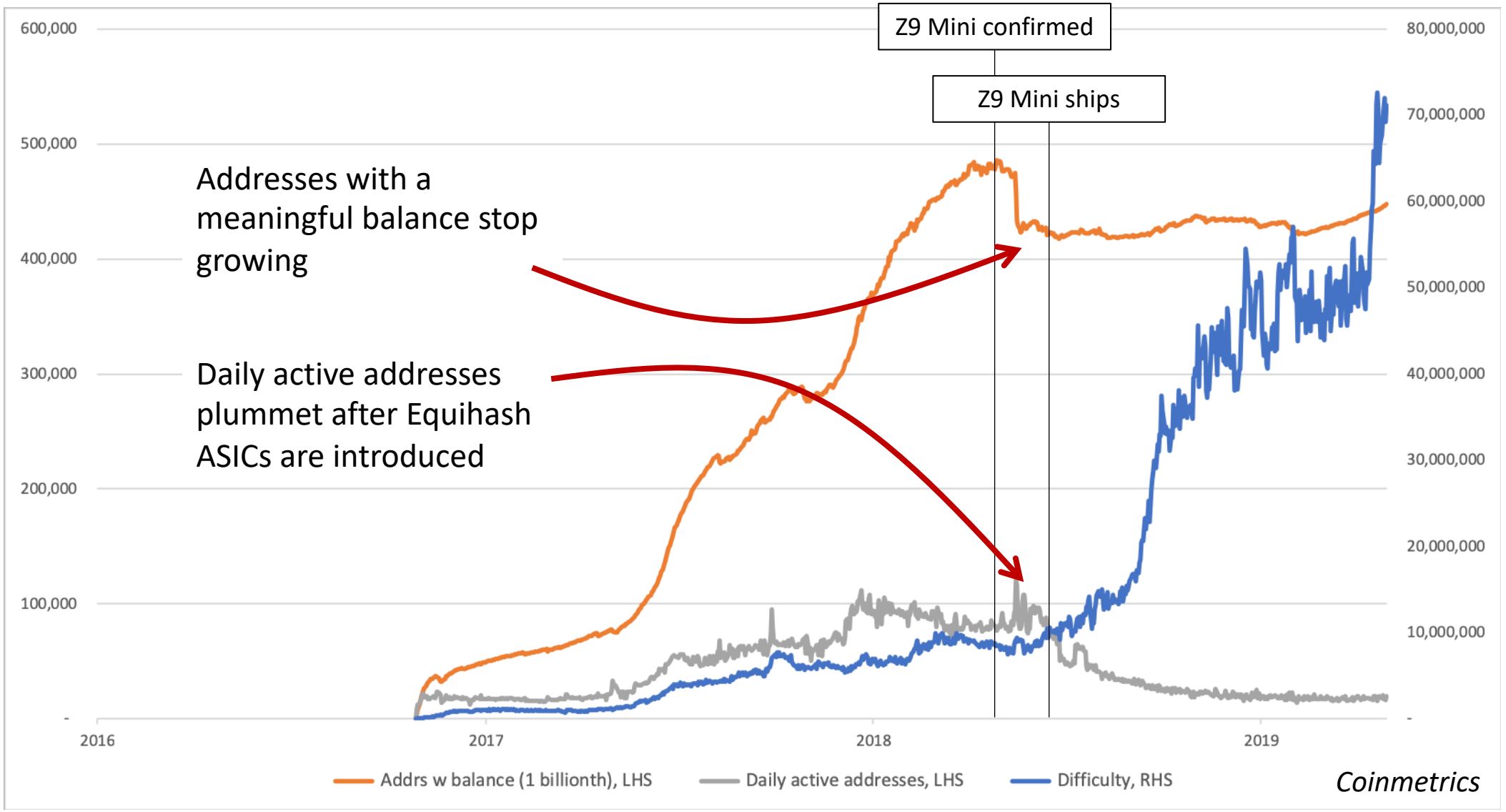


Coinbase Earn airdrop contributed to a gain of 32,000 t-addrs with at least 0.01 ZEC

Evaluating the 0x Airdrop

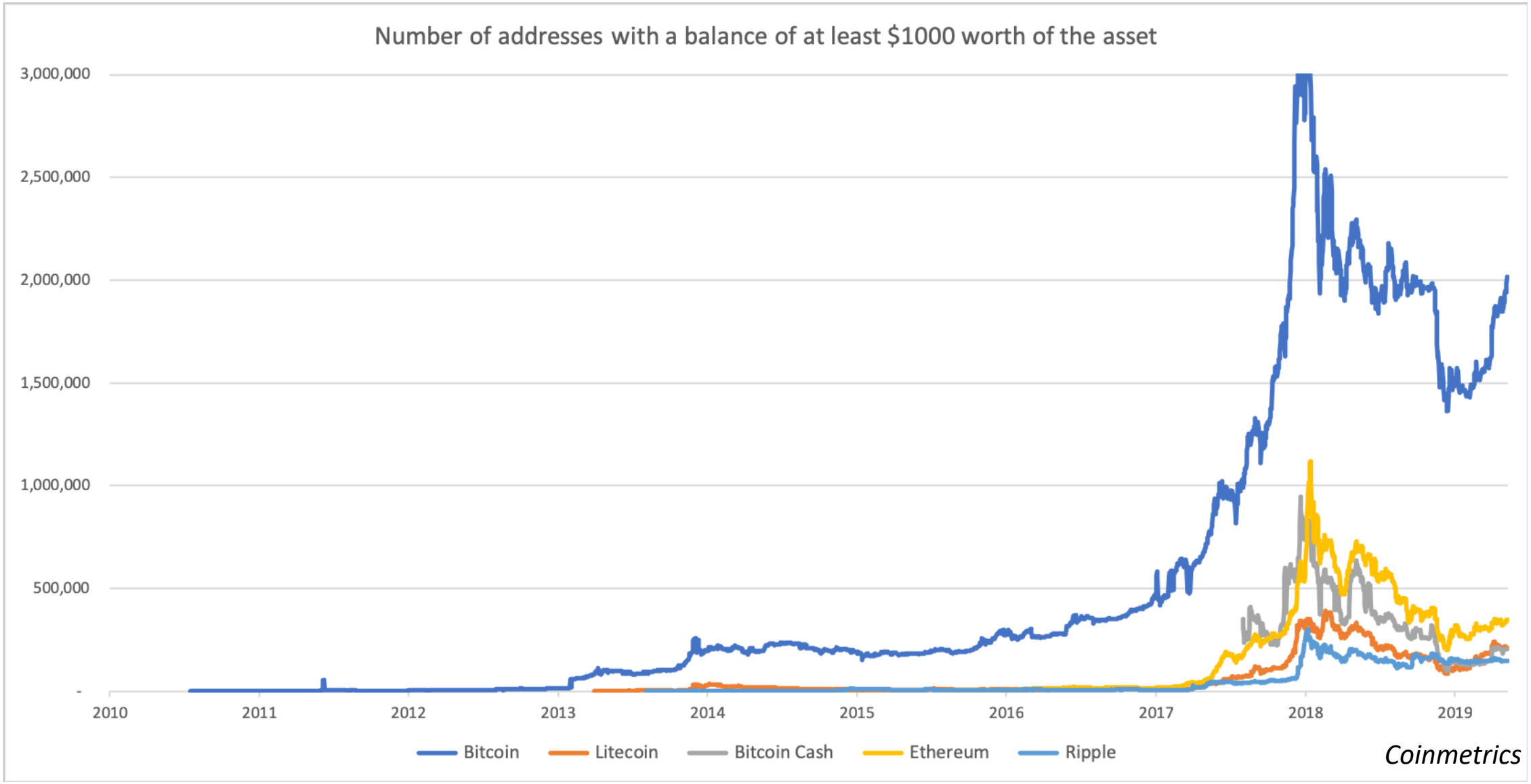
While 100,000 users may have claimed the airdrop, only at most 10,000 recognized the tokens on-chain

Measuring network health: the effect of ASICs on Zcash



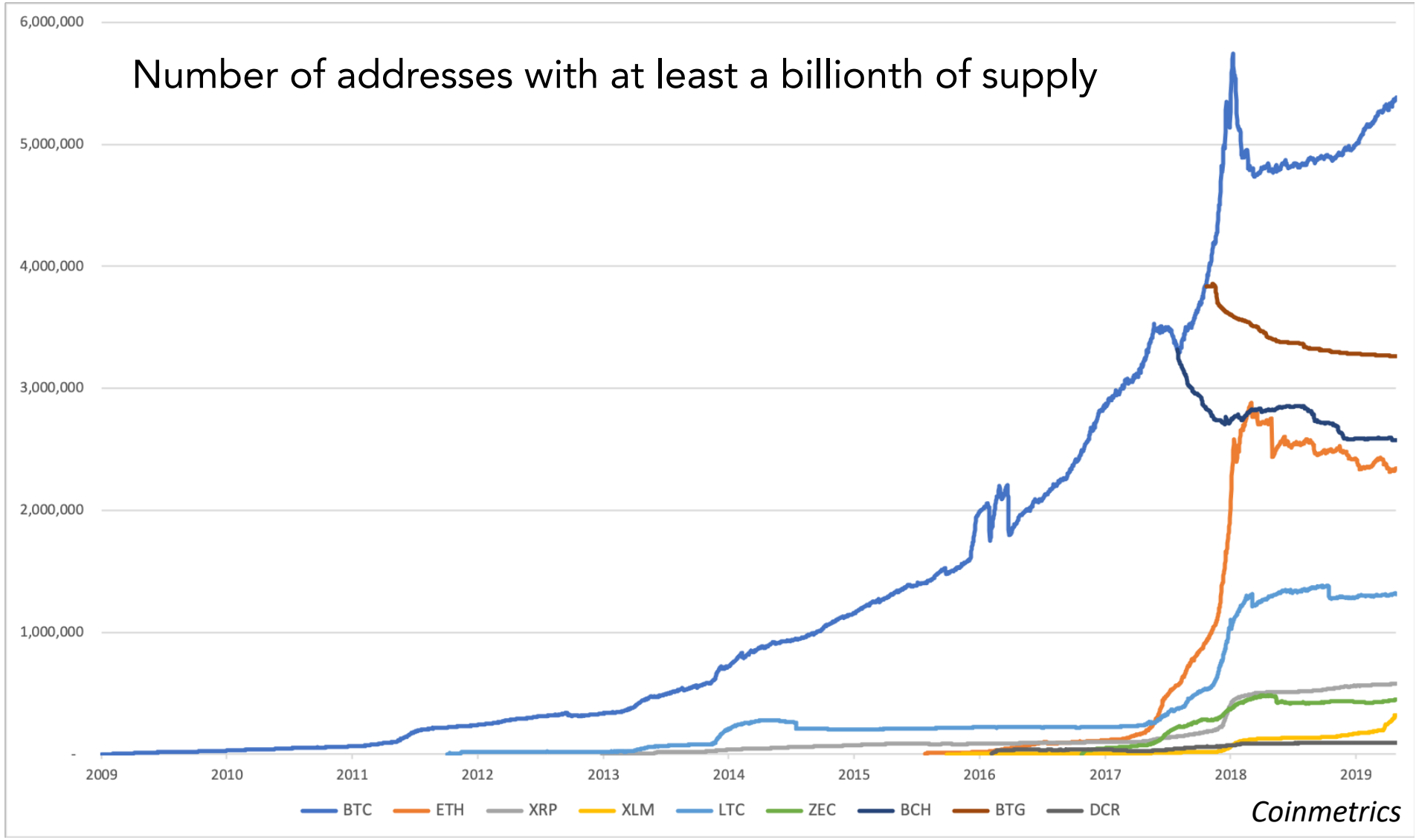
Early evidence that ASICs may reduce network vibrancy; first pointed out by Brian Venturo of Atlantic Crypto

Assessing relative network vibrancy



Coinmetrics

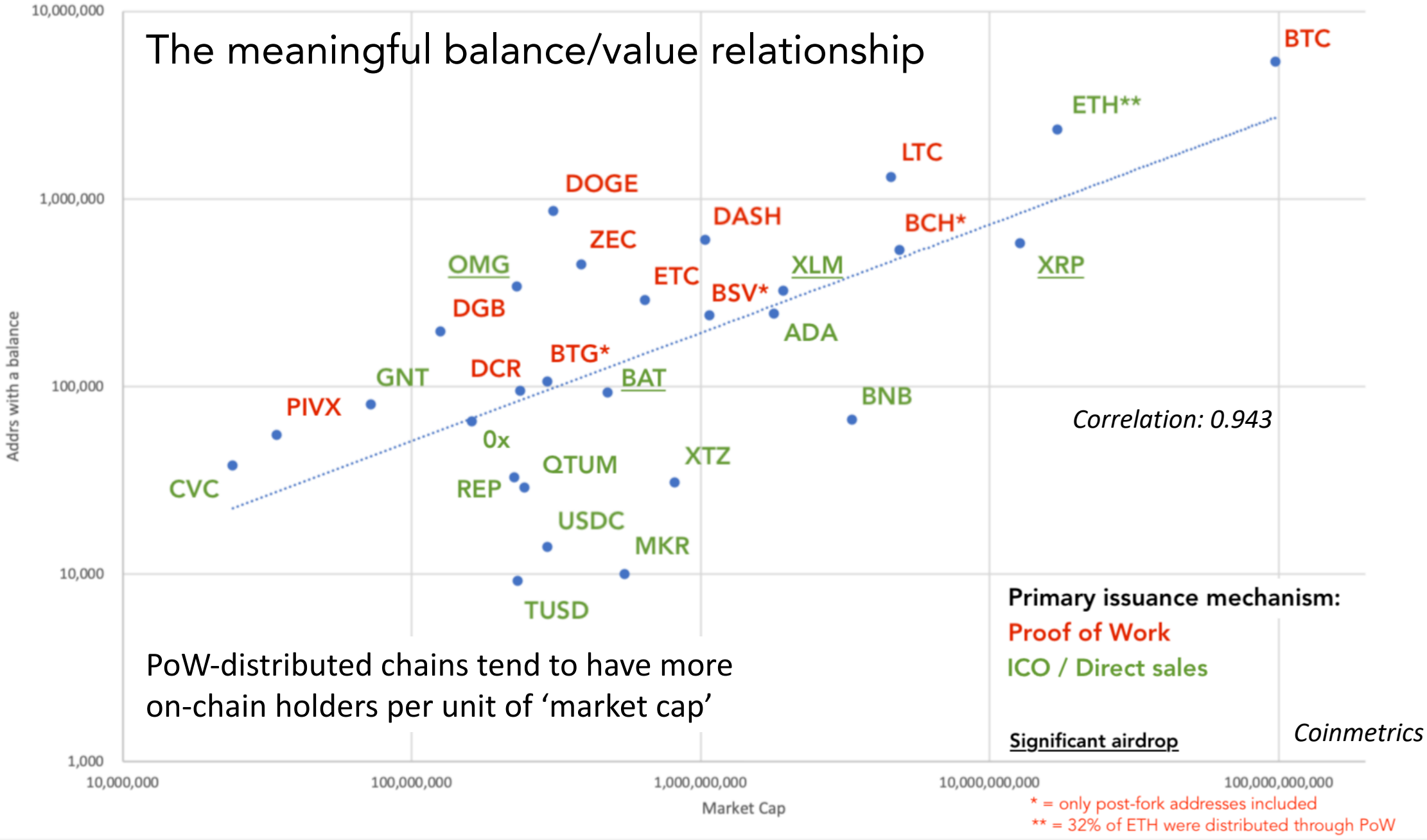
Assessing relative network vibrancy: dispersion



This approach rules out dust addresses with a minuscule balance

Mappings between account based and UTXO based chain aren't perfect

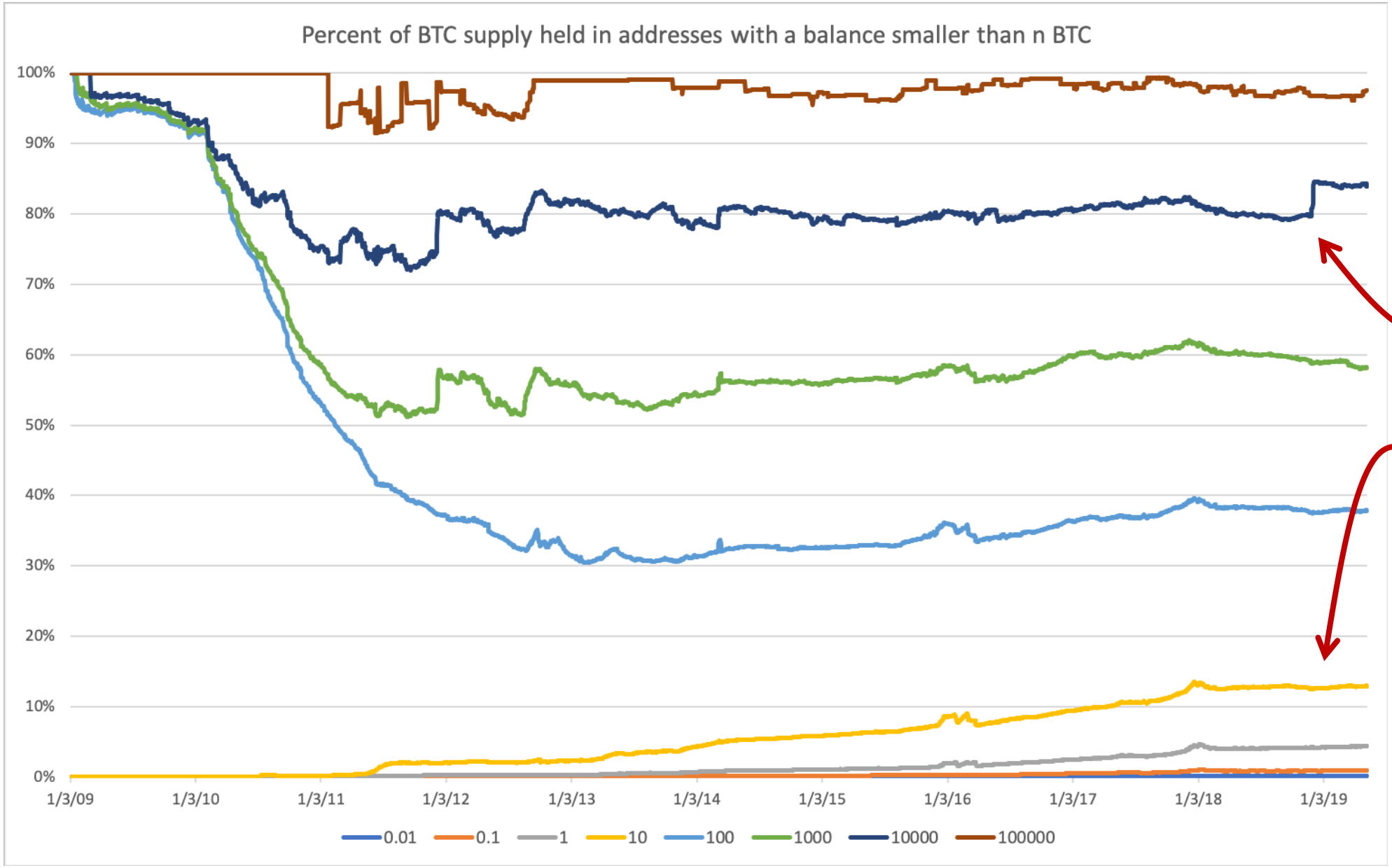
The meaningful balance/value relationship



PoW-distributed chains tend to have more on-chain holders per unit of ‘market cap’

Coinmetrics

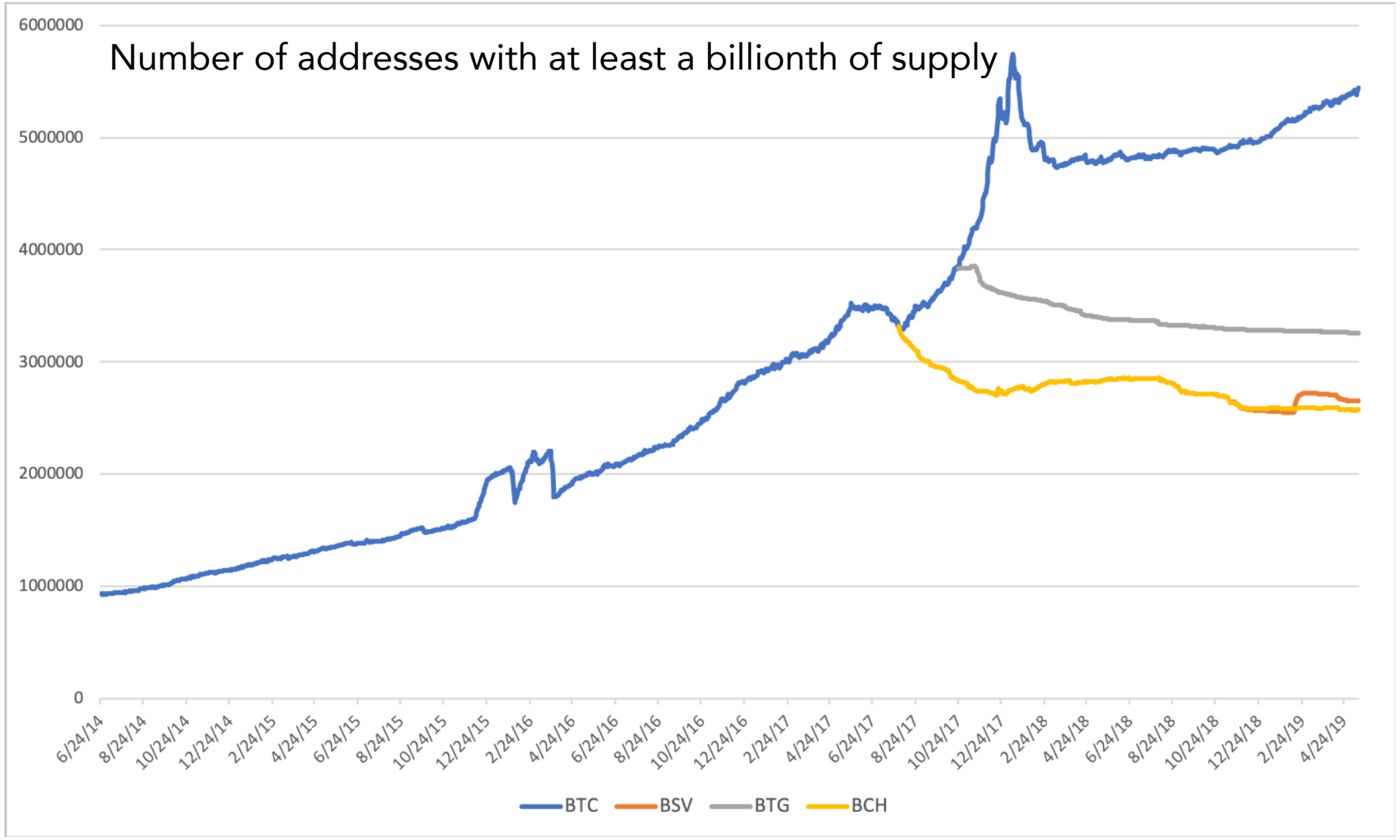
Assessing dispersion: supply repartition by balance



Coinbase reshuffles cold storage; moves coins from wallets of 60,000 BTC to <10,000 BTC

Small wallets represent a growing fraction of BTC supply; BTC is gradually becoming more dispersed and less dominated by whales

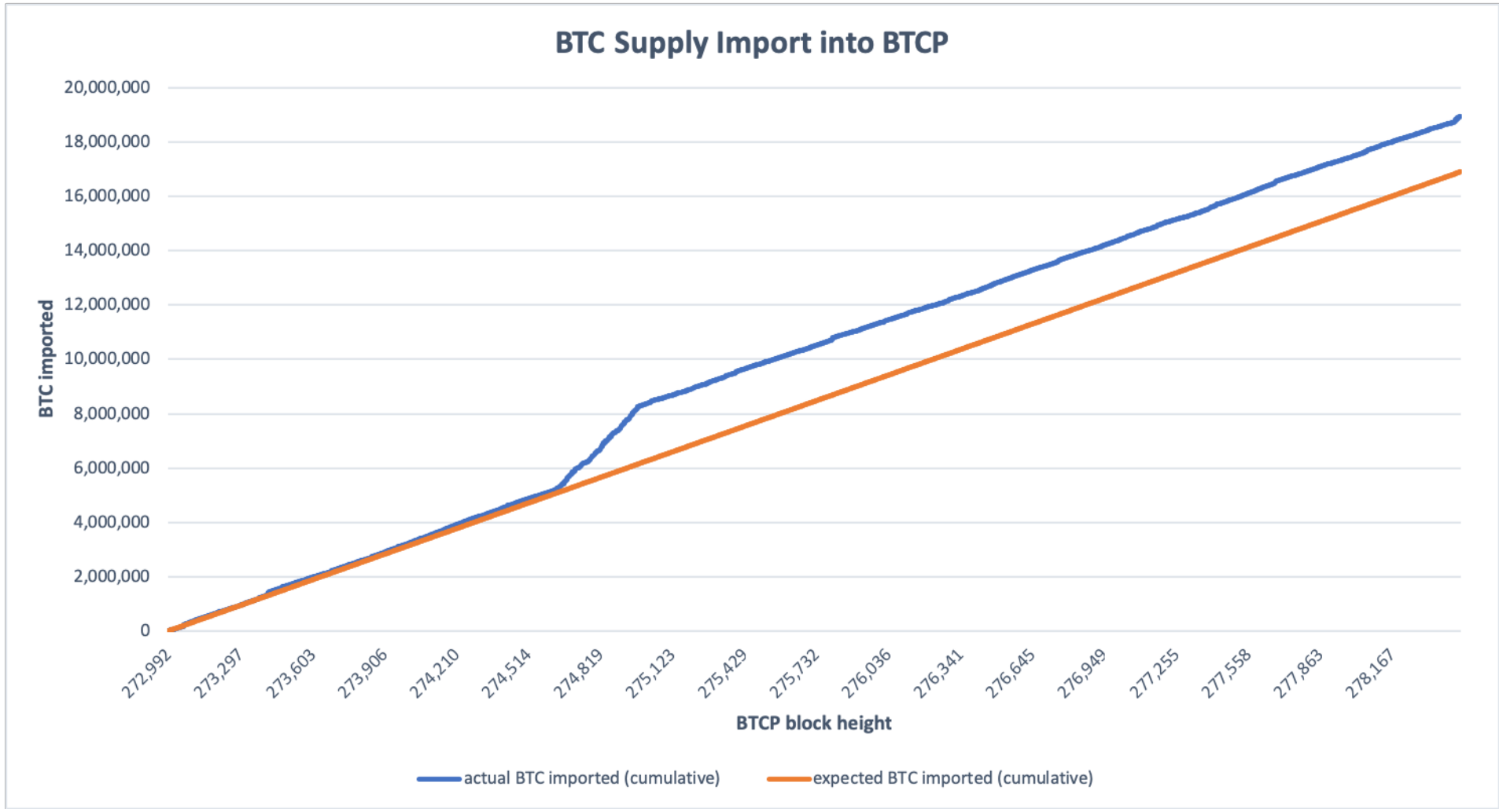
The concentrative effect of forks



All Bitcoin forks have effectively had a concentrative effect as small holders sold and whales scooped up supply

If the value of the network is the dispersion of the UTXO set, all forks have been failures so far

Ensuring the integrity of the chain: BTCP case study



During a routine supply audit of Bitcoin Private, we discovered that an additional 10% of supply had been covertly minted during the UTXO import

Ensuring the integrity of the chain: XRP case study

Quarter	XRP released – returned (reported)	XRP released – returned (on-chain)
Q1 2018	3.0B – 2.7B = 0.3B	3.0B – 2.7B = 0.3B
Q2 2018	3.0B – 2.7B = 0.3B	3.0B – 2.7B = 0.3B
Q3 2018	**3.0B – 2.6B = 0.4B**	**3.0B – 2.5B = 0.5B**
Q4 2018	3.0B – 2.4B = 0.6B	3.0B – 2.4B = 0.6B
Q1 2019	**3.0B – 2.3B = 0.7B**	**3.0B – 2.2B = 0.8B**

- In vetting Ripple's quarterly disclosures regarding their escrow releases, we found that they under-reported the XRP released from escrow by 200m XRP (~$76m at today's prices)

- This in effect constituted additional dilution not reported to investors and users

Future directions for on-chain data

- Provable solvency for custodians and exchanges, provable collateralization ratios
- Granular macroeconomic data: real-time GDP, inflation, interest rates, etc
- Fully transparent financials for foundations, nonprofits, etc
- On-chain cashflows for new forms of organizations will enable instant disclosure rather than quarterly reports – markets will be able to efficiently price in new information

Where does crypto data go from here?

On-chain data

- Standardize definitions & methodology
- Acknowledge that many metrics aren't apples to apples (txn count, eg)
- Acknowledge the financial incentive to generate misleading data
- Impose robust taxonomies

Exchange data

- Whitelist, don't blacklist
- Put the burden on exchanges to demonstrate that their data has integrity
- Don't reward providers that are naïve or don't vet exchange data
- Make skepticism the default
- Reward exchanges and providers that adhere to conventional data standards

Both

- Aim for consistency, standardization, and avoid motivated reasoning
- Be aware of goodhart's law at all times: when a measure becomes a target, it ceases to be a good measure