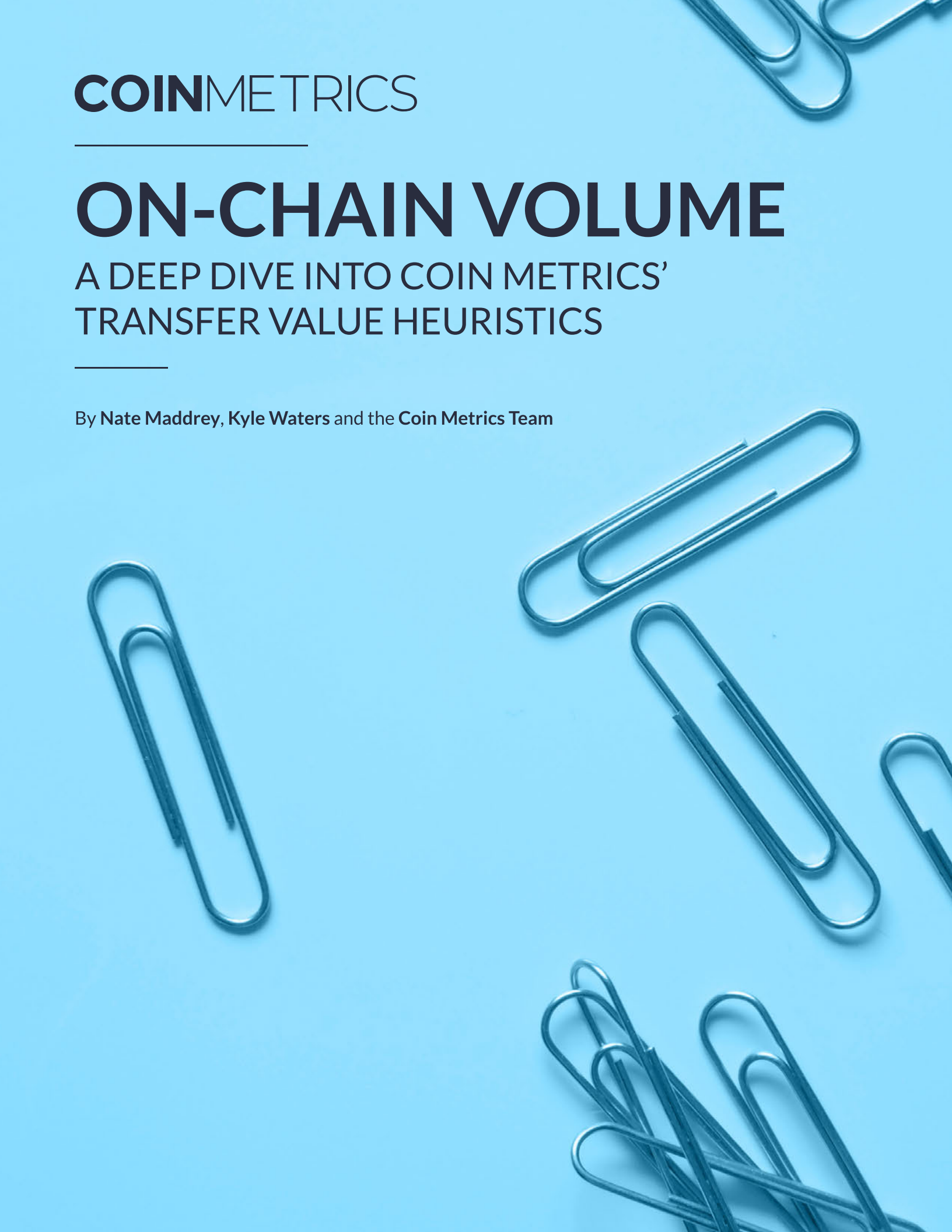
COINMETRICS

ON-CHAIN VOLUME

A DEEP DIVE INTO COIN METRICS' TRANSFER VALUE HEURISTICS

By **Nate Maddrey**, **Kyle Waters** and the **Coin Metrics Team**



A DEEP DIVE INTO COIN METRICS'

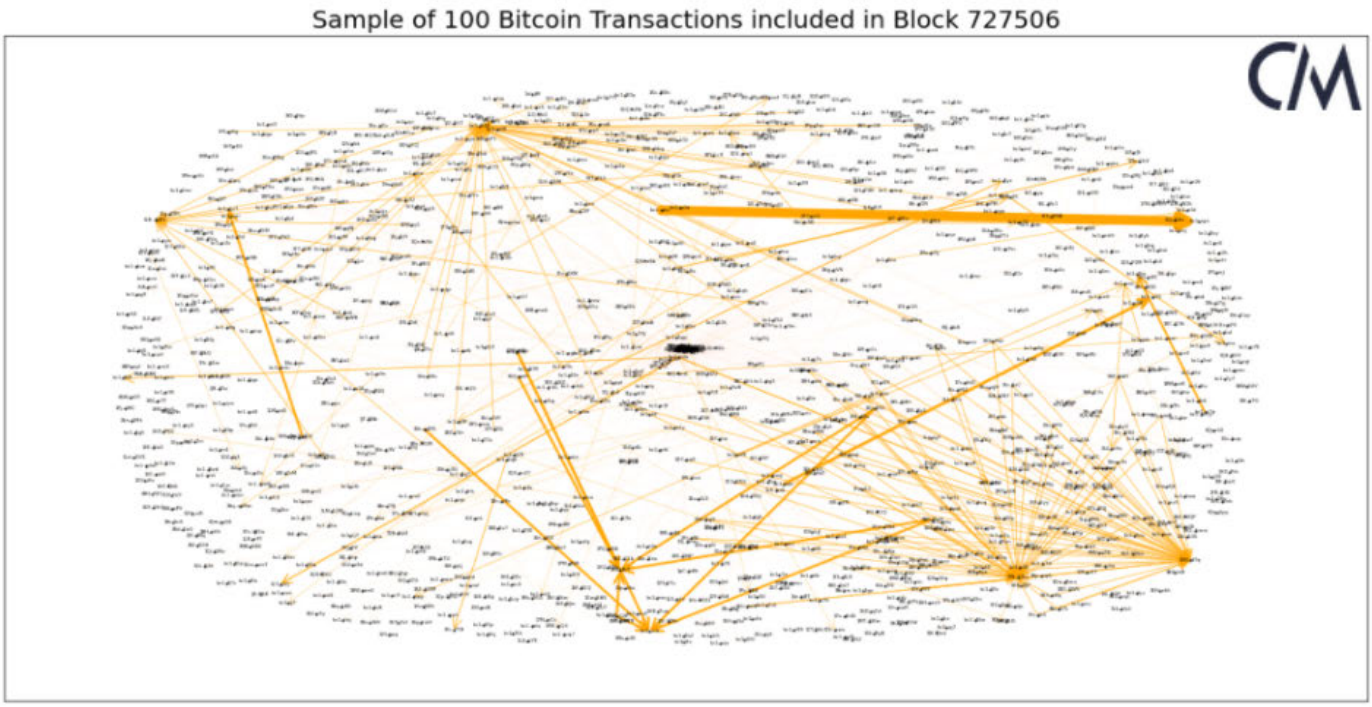
TRANSFER VALUE HEURISTICS

Public blockchains are settling more value than ever before. But calculating on-chain transfer volume is a non-trivial exercise. Raw data is rife with noise and opportunity for misinterpretation. To create better estimates, Coin Metrics implements a series of rules or "heuristics" to create adjusted on-chain volume metrics. In this report, we walk through our transfer value heuristics using real-world examples to give intuition for their effectiveness.

A DEEP DIVE INTO COIN METRICS'
TRANSFER VALUE HEURISTICS

In 2021, more value was transferred across Bitcoin and Ethereum in dollar terms than ever before, with both networks settling amounts measured in the trillions. This is exactly measurable because public blockchains are distributed and open databases that allow anyone, anywhere in the world to access the entire history of transactions and amounts transferred between users.

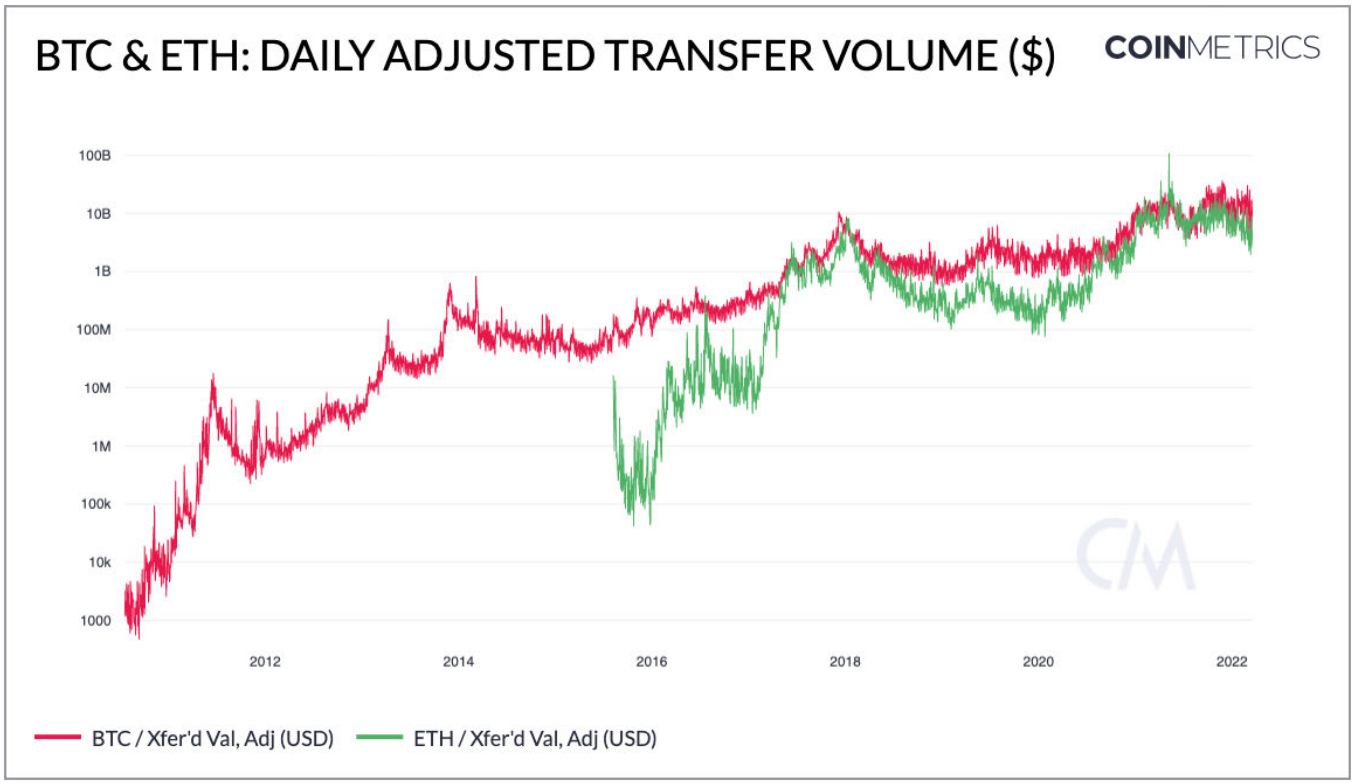
Public blockchains are unprecedentedly rich economic datasets, but using blockchain data ("on-chain" data) for economic analysis is a fragile exercise. Raw on-chain data is rife with noise and ample opportunities for misinterpretation. Yet, as crypto adoption accelerates and policymakers consider the implications of public blockchains, it is crucial to contextualize and rigorously examine their data and growth.



Flow of bitcoin in 100 sample transactions extracted from Bitcoin block 727506, data compiled from Coin Metrics' ATLAS.

Coin Metrics was founded out of this exact need for high quality, carefully-curated crypto data. Some of Coin Metrics' earliest research (c. 2018) outlined the arduous task of estimating on-chain volume.

Today, Coin Metrics approaches this task by applying a series of rules or "heuristics" to the raw data, informed by leading industry research and our own fundamental understanding of blockchain data. These rules ultimately help us siphon out noise and better determine the significance of burgeoning public blockchain economies. We currently offer our adjusted transfer volume estimates across 116 unique assets to both community and pro CM users alike.



Source: *Coin Metrics' Network Data Charts*

In this special report, we break down the complicated effort of estimating public blockchains' economic throughput with the use of heuristics.

THE DIFFICULTIES OF BLOCKCHAIN DATA

It is vital to understand the sources of difficulty in accurately gauging on-chain transfer volume. They can be broken down into two buckets: adjustments needed for non-meaningful activity and adjustments arising from blockchain design.

Non-meaningful activity comes in many different flavors. While the designation of non-meaningful activity[1] is somewhat subjective, a general theme is that the owner of the asset does not change during the transfer. Here are a few examples.

- **Self-churn:** transfers with the same sender and recipient (i.e. with yourself)
- **Consolidation:** another form of self-churn, where users aggregate funds from multiple addresses into a single address for ease of management or privacy, especially important for UTXO-based chains like Bitcoin[2]
- **Exchanges:** custodial exchanges will often move funds around to meet demands for customer withdrawals or to improve cold wallet security
- **Mixers:** services that move funds around in an attempt to obscure their origins
- **Spam:** junk transfers, perhaps to game network statistics (e.g. wash trading), pass along messages or advertisements on-chain, or worse, from malicious actors trying to clog the network
- **Other Pass Throughs:** some transfers are simply intermediary transactions, an example being transactions triggered via automated smart contracts that can hold and send funds

It's important to note that perfectly segmenting activity like the above is futile because it would require foolproof tagging of addresses associated with such behavior. However, as we explain later below, there are some simple yet powerful filters to make effective adjustments.

[1] It's worth considering that from the perspective of the blockchain's security, arguably no transaction is economically meaningless. Miners still receive a subsidy and the transaction contributes to the security of the network.

[2] See this article from Casa, a Bitcoin wallet provider.

The other challenge arises from different blockchain design schemes. Today, there are two major blockchain design philosophies: unspent transaction output (UTXO) based and account-based. Understanding how each works is essential to developing and interpreting appropriate heuristics. Here are some simplified and brief analogies describing each.

- **UTXO-based:** each coin is uniquely referenced and tracked on-chain. UTXO-based chains are conceptually similar to physical wallets that are composed of individual bills and coins - although a wallet may collectively hold $100, it could be composed of individual $10 or $1 bills, each of which is individually tracked. Imagine controlling paper bills of various denominations: $5, $10, and $20. When you go to purchase a good for $14 you can either use the combination of the $5 and $10 bills, and receive $1 in change, or use the $20 bill and receive $6 in change. As with physical paper bills, the UTXO model requires spending full amounts and receiving change back. Examples include Bitcoin, and Bitcoin forks.

- **Account-based:** there is no unique trace of each individual coin. Rather, coins are tracked as balances in accounts, more akin to traditional bank accounts. When transactions occur, the system's "state," or list of everyone's balances, changes to reflect new activity. Examples include Ethereum and most other smart contract platforms.

Given the structural differences between UTXO and Account-based blockchain design, different approaches must be applied to each. The table below summarizes the heuristics that Coin Metrics currently uses when calculating our base transfer value figure (TxTfrVal) and adjusted transfer value (TxTfrValAdj), for both native and USD denominated data.

Coin Metrics' Transfer Value Heuristics

Type	Heuristic	Transfer Value (TxTfrVal)	Adj. Transfer Value (TxTfrValAdj)
UTXO	Obvious Change/ Self-Churn	✓	✓
	Early Spend		✓
	Round Numbers		✓
Account	Self-Churn	✓	✓
	Pass Through		✓

As of March 2022.

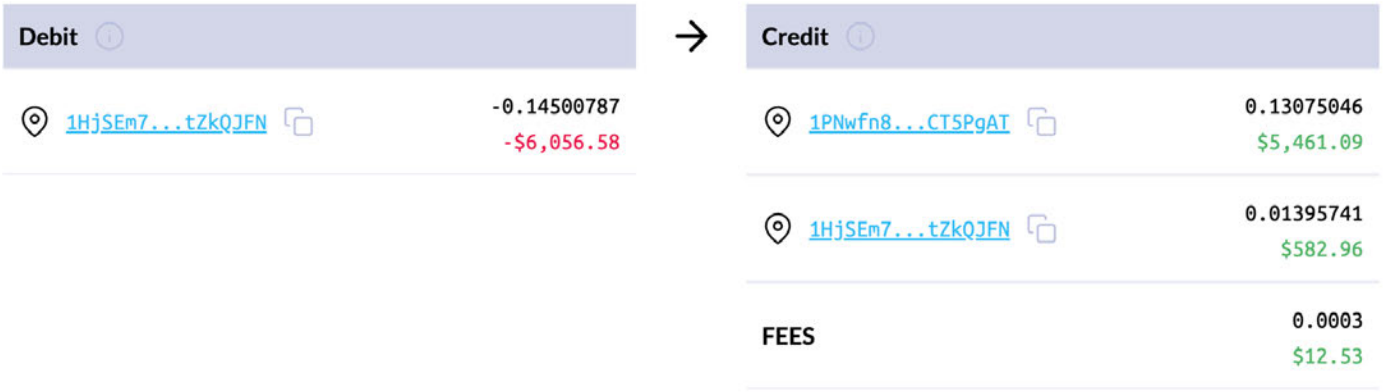
Note: coinbase transfers (i.e. issuance) are also excluded from all transfer value figures.

In the next section, we provide some intuition for each using real-world examples gathered on-chain.

UTXO-based Heuristics

Obvious Change

As described above, UTXO chains operate like typical cash transactions, often involving change. To illustrate, take this example Bitcoin transaction included in block 552399.



Source: *Coin Metrics' ATLAS*

In the transaction above, the Bitcoin address 1HjSEm7...tZkQJFN sends a total of 0.145 BTC (more specifically .14500787 BTC, but we've rounded down for easier readability). Most of this is sent to the address 1PNwfn8...CT5PgAT, which receives 0.131 BTC, while the remainder flows back to the *original sender* as change (with a small amount allocated to miner fees). This example occurred because the sender wanted to send .131 BTC but not own a set of UTXOs that equaled that exact amount, so needed to send a little more than required receiving change in return.

A naive approach to measuring transfer volume in this example would simply take the full 0.145 BTC as the value transmitted. However, it's clear that the output of 0.014 BTC to 1HjSEm7...tZkQJFN is simply change and recycled back to its previous owner. To account for this, we employ a heuristic known as *obvious change* that removes outputs flowing back to the originating address.

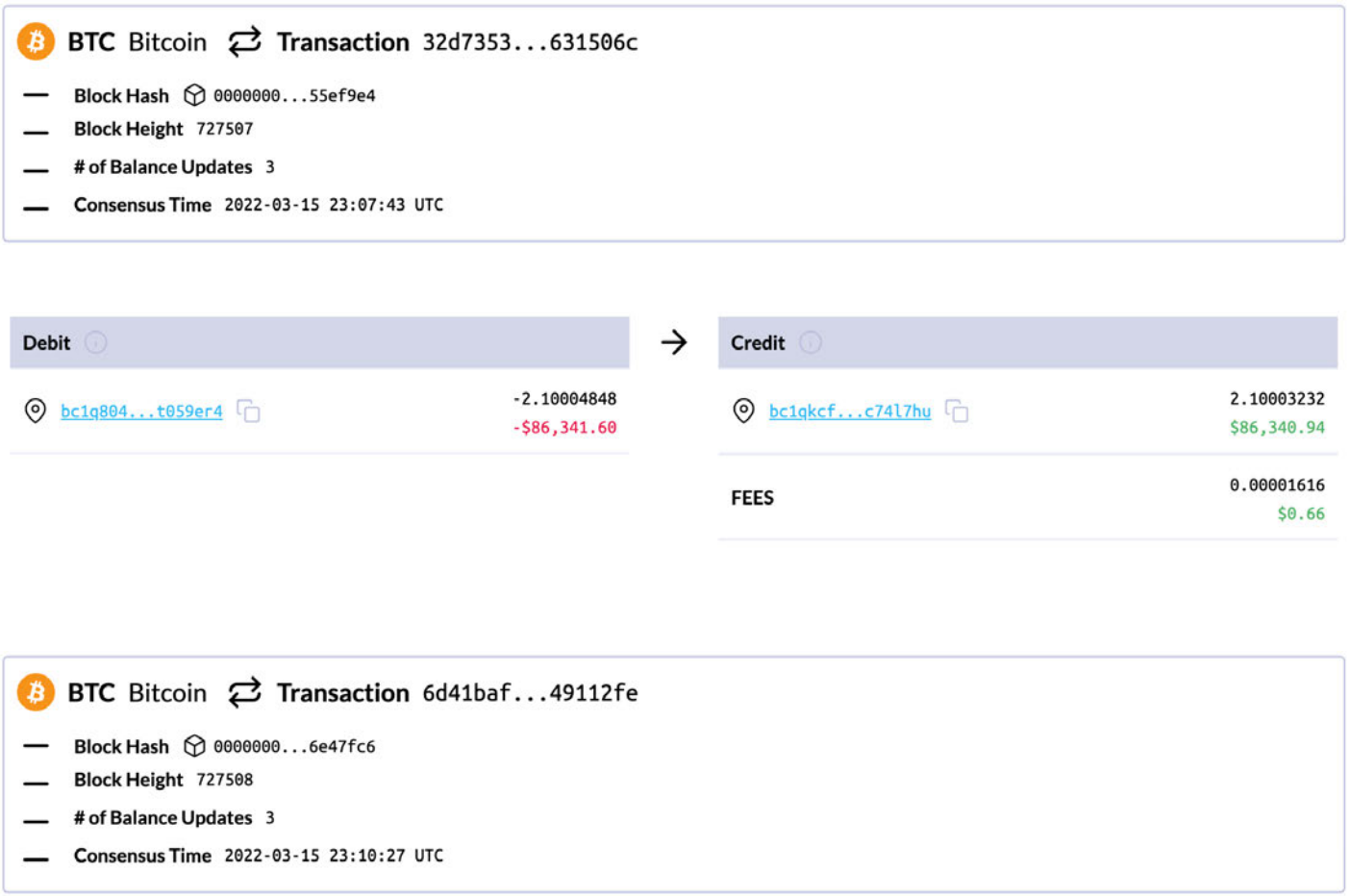
Left uncorrected, obvious change outputs overstate transfer value by a factor of 5x or even 10x more versus when they are removed. For this reason, both our baseline and adjusted transfer value figures remove obvious change from UTXO chains.

Early Spend

From exchange activity to mixers to spam, non-meaningful transfers can take many different forms. Tagging each of these entities' ever-changing sets of addresses on-chain (which can measure in the tens of millions for a single exchange alone) is a daunting task to keep up with. Thankfully, there are some useful heuristics that can help discount this activity.

One commonality of transfers involving mixers, exchanges, and the like is that they tend to have abnormally high velocity. Often, funds will bounce between addresses in a short amount of time to reach their final destination. Comparatively, regular payments are usually not spent right away. Using this observation, the *early spend* heuristic was born. First introduced in a 2017 paper by Naranayan et al., the early spend heuristic recommends discarding transfers where the sender re-sends the funds to another address shortly after receiving them. In Coin Metrics' implementation, we found it most appropriate to use a time cut off of one hour (roughly within six blocks for Bitcoin).

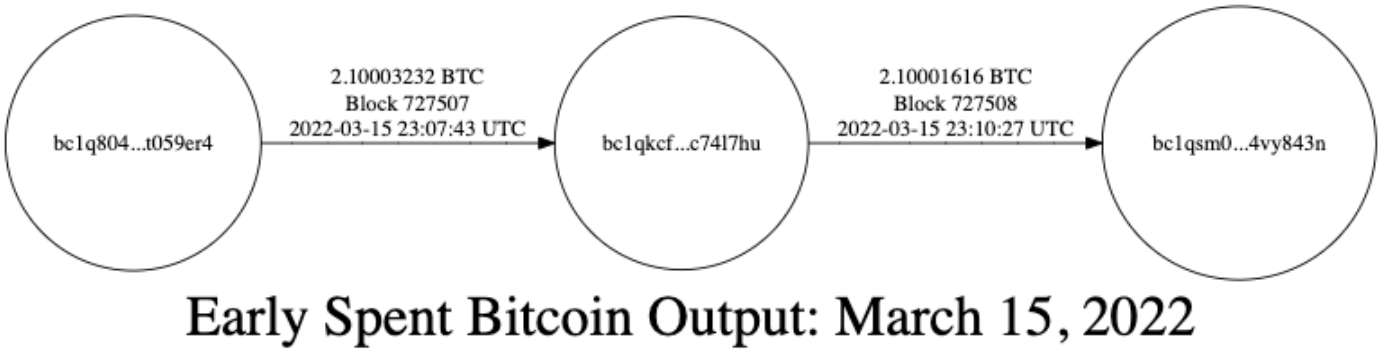
Below is an example of a case where we use the early spend heuristic. In this first transaction, bc1q804...t059er4 sends ~2.1 BTC to bc1qkcf...c74l7hu, at block height 727507 (11:07 PM UTC). Then, just one block later at 11:10 PM, the output of the previous transaction is spent in another transaction and sent to bc1qsm0...4vy843n.



Debit		Credit	
bc1qkcf...c74l7hu	-2.10003232 -$86,403.73	bc1qsm0...4vy843n	2.10001616 $86,403.06
		FEES	0.00001616 $0.66

Source: *Coin Metrics' ATLAS*

Viewed as a graph below, the output spent in block 727508 looks like a pass through. While it's unclear exactly why the BTC was moved so quickly, it's unlikely a real ownership change occurred from block 727507 to 727508.



Early Spent Bitcoin Output: March 15, 2022

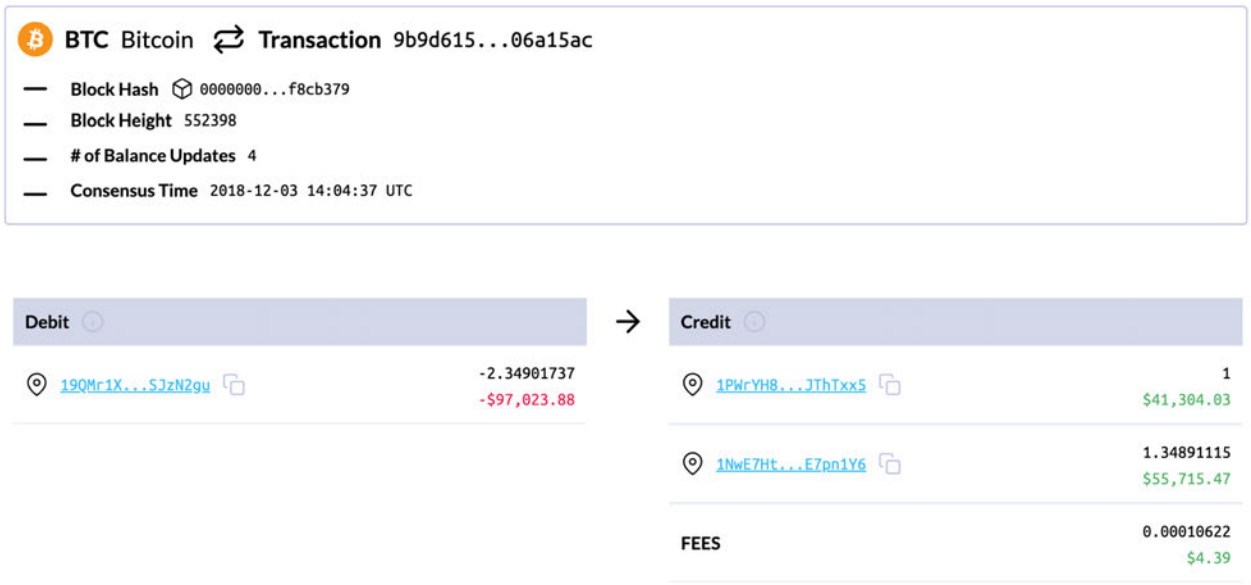
One possible explanation of this type of behavior is exchange-related activity. Centralized exchanges drive significant amounts of activity on public blockchains today and custody crypto assets on behalf of millions of people. According to Coin Metrics' data of March 19, 2022, at least 8% of all BTC supply (1.5M) was held on exchanges. Exchanges will often move funds around internally to conduct day-to-day operations or to consolidate holdings for security or efficiency gains.

Whatever their purpose, Naranayan et al. also inspected transactions like the ones above when first adopting the early spend heuristic and concluded that "manual examination suggests that such transactions are highly likely to represent self-churn, such as 'peeling chains' where a large output is broken down into a series of smaller outputs in a sequence of transactions."

Round Numbers (“Assumed Change”)

Economics is ultimately a social science and the study of public blockchain economies is no different. In everyday commerce, humans often tend to anchor to and price goods and services in convenient, easy-to-remember numbers. Building off of this observation, we can make reasonable assumptions about the patterns of “real” exchange that occur on-chain. One rule we implement is to explicitly search for *round-numbered* outputs and assume that outputs with extra precision are change.

In this example transaction below, 19QMr1X...SJzN2gu spends a UTXO of ~2.35 BTC, of which exactly 1 BTC is outputted to 1PWrYH8...JThTxx5 and 1.34891115 to 1NwE7Ht...E7pn1Y6.



Source: *Coin Metrics’ ATLAS*

With the round numbers heuristic, we decide to count the exact output of 1 BTC as real economic activity and discard the output with extra digits, as it is likely to be change.[3]

[3] It’s possible that some high-precision amounts, when translated to US dollar terms, might actually be round numbers. However, this type of heuristic could be prone to error due to changes in pricing from the point of the transaction entering the mempool to its inclusion in a block. Moreover, there might be many more reference points than just US dollars, e.g. euros, yen.

As mentioned earlier, identifying change outputs is a very important task when gauging transfer value in UTXO-based blockchains. It's important to note that in most situations, address reuse is what leads to non-obvious change outputs. If users are generating new addresses for change outputs, it is more difficult to determine if it is self-churn or not. But recent data suggest that about half of all Bitcoin outputs today re-use addresses.

However, there are shortcomings to focusing too much on change detection. For example, most research focuses on transactions with 2 outputs, one of which is always change. While this pattern was most common in Bitcoin's early days, now batching (many outputs in a single transaction) and change elimination both potentially render this assumption useless. As such, it's important to design heuristics that can survive paradigm changes.

ACCOUNT-BASED HEURISTICS

Our approach to account-based transfer volume heuristics builds upon the insights of UTXO-based heuristics while taking into account the nuances of their different design mechanisms.

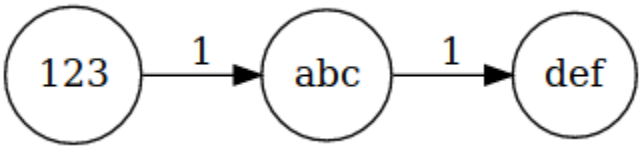
Self Churn

As is the case with UTXO blockchains, account-based blockchains may also feature transfers where the sender and receiver are the same address. We remove all such volume from both our headline transfer volume and adjusted transfer volume data.

Pass Throughs

In the case of UTXO-based models, the early spend heuristic is a powerful rule to eliminate non-meaningful intermediary activity. However, because each coin in an account-based blockchain lacks a unique timestamp and reference, the early spend heuristic is not possible to directly implement. But a better estimate of transfer volume still requires filtering out pass throughs. To solve this, we developed a filtering technique that maps the logic of the early spend heuristic to the universe of account-based blockchains.

One of the most common cases of pass through activity on account-based chains is exchange deposit addresses. When a user wants to deposit 1 token onto an exchange, the exchange will generate a deposit address (0xabc), request the user send the token to that address from one of their own accounts (0x123), and immediately move that deposit in another exchange-controlled hot wallet for that token (0xdef):

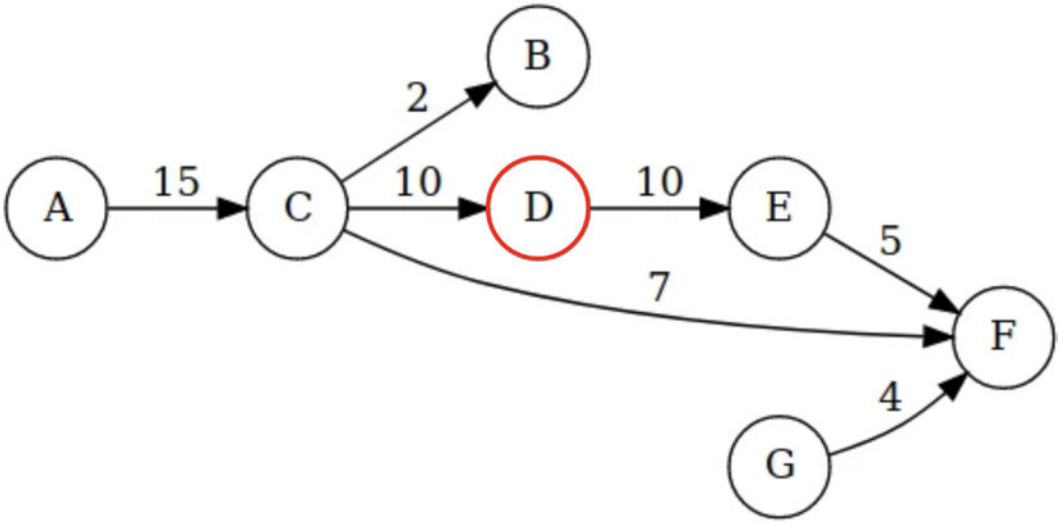


Exchange Deposit Addresses as Pass Through Activity

In this case, the naive transacted volume is 2, but in reality, it should be considered to be 1 since those transactions happen in quick sequence: 0xabc is just a passthrough account. Eliminating the middle transaction is similar to using the UTXO based early spend heuristic.

The heuristic works by collecting all transfers within the past hour and finding the amounts exchanged between accounts. If an account has received and sent an equal amount in the last hour (i.e. incoming and outgoing transfers netting out to zero), it is deemed a pass through account and ignored.

The network graph below presents a hypothetical example where each circle (node) in the graph represents an account and the flow between accounts are amounts transferred.



Filtering Technique to Identify Pass Through Volume, Hypothetical Hour of Transfers

The transferred volume is taken as the total amount of funds received by accounts not deemed to be pass throughs. Looking at the graph above, account D is an obvious example of a pass through, with 10 received from C and subsequently sent to E within the hour. Following the heuristic, the received amount of 10 is **excluded**. Thus, the total heuristic-adjusted transfer value in this mini account-based blockchain economy would be the total amounts received by each account: C (15) + B (2) + E (10) + F (16) = 43.

EVALUATING HEURISTIC PERFORMANCE

Without the ground truth, it's impossible to exactly measure the performance of transfer value heuristics. However, there are methods to establish confidence in the results.

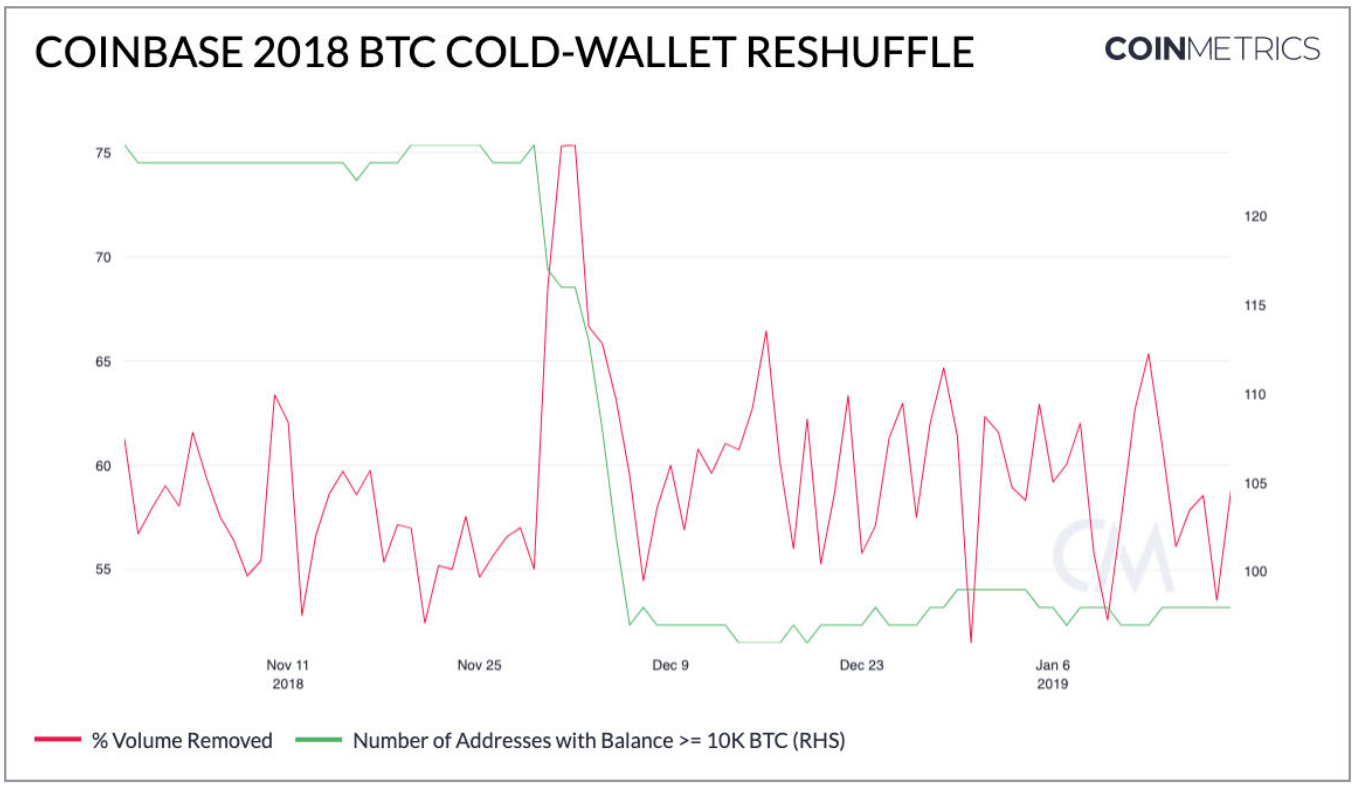
First, studying the heuristics' performance during specific on-chain events in the past illuminates their credibility. Known exchange activity provides a good case study. It is well known that centralized exchanges are often the source of anomalous spikes in on-chain transfer volume when conducting reshuffles of crypto assets held in cold storage.[4] Not all cold-wallet reshuffles are publicly broadcasted, however, there are some particularly large reshuffles in the past that have been confirmed by the exchange.

One such example is Coinbase's December 2018 cold-wallet reshuffle, described by Coinbase as the "the biggest (and quietest) crypto transfer on record" in which 5% of all BTC, 8% of all ETH, and 25% of all LTC in circulation were moved. The impact of this movement is clearly visible on-chain.

The chart below shows the effectiveness of our UTXO-based heuristics in filtering out the cold-wallet reshuffle from our adjusted transfer volume. The line in green shows the effect of the reshuffle on Bitcoin address balances, plotting the number of Bitcoin addresses with more than 10K BTC, which fell precipitously in a few days at the end of November 2018. The reshuffle saw the exchange move from a few large wallets to a more distributed set of wallets.

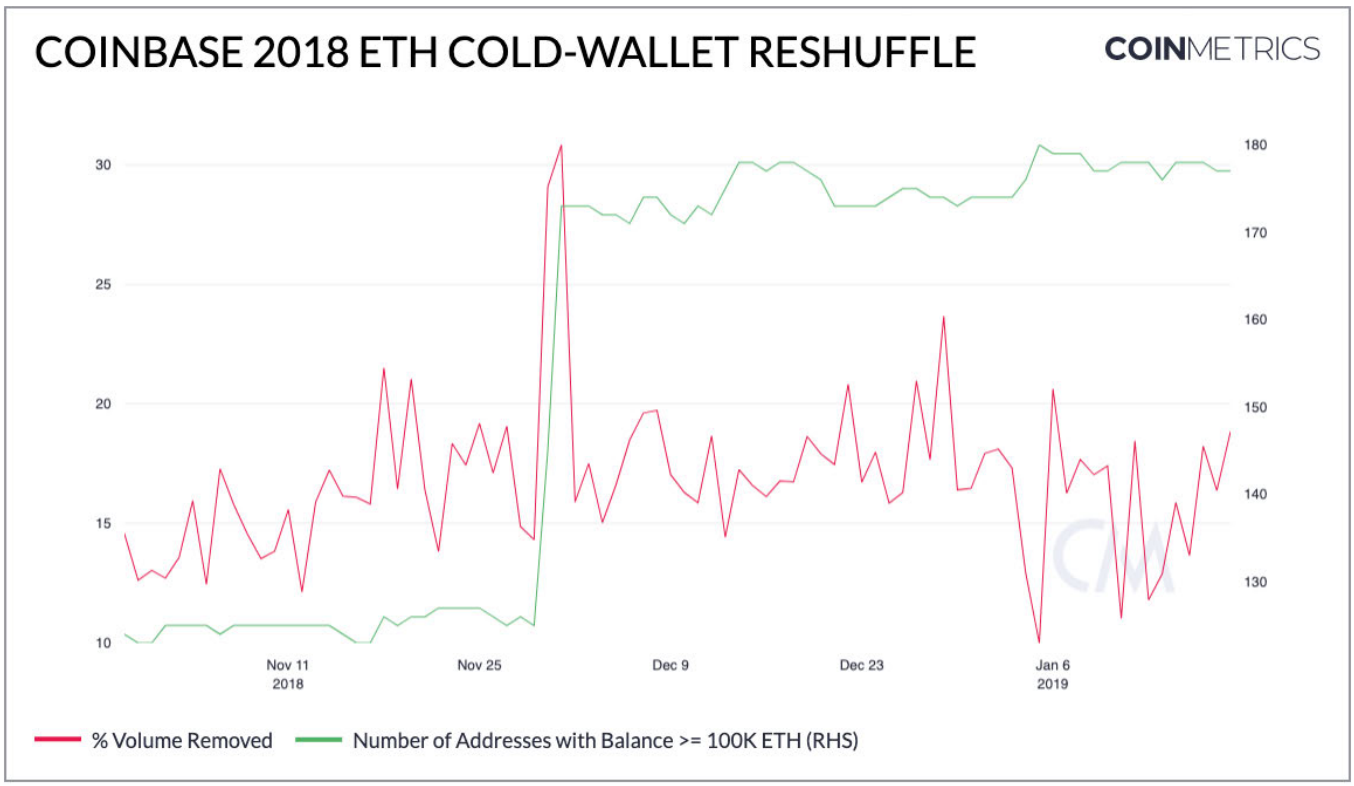
Plotted in red is the percentage of BTC transfer volume removed from our baseline figure by day. During the reshuffling, the percentage of volume removed jumped from the 55-60% range to 75%. This is because more volume was captured by our heuristics that filter for self-churn, early spends, and rounded change outputs.

[4] Assets held in wallets whose private keys are stored offline, often with purpose-built physical devices, and entirely disconnected from automated systems and the web.



Source: *Coin Metrics' Formula Builder*

A similar effect is visible for ETH (though the reshuffle seemed to consolidate rather than redistribute funds).



Source: *Coin Metrics' Formula Builder*

Finally, studying unadjusted total value transferred against adjusted totals by crypto asset yields valuable insight into the heuristics' effectiveness. The table below shows the total value transferred in 2021 using our baseline metric and heuristic-adjusted total. The column to the far right shows the impact of adjustments, measured by the percentage of volume removed when moving from the baseline to adjusted figure.

For example, using our baseline calculation Bitcoin settled $13.1T in 2021 and $4.7T on an adjusted basis. Ethereum, meanwhile, settled $5.4T and $3.6T (in ETH alone), unadjusted and adjusted.[5] To put these numbers in context, credit-card giant Visa reported total payments volume of $10.4T in 2021.[6]

[5] Including ERC-20 token transfers adds an additional ~$8T in adjusted transfer volume.

[6] This is not a perfect comparison of value-transfer systems, but offers some perspective nonetheless.

2021 Transfer Volume by Asset

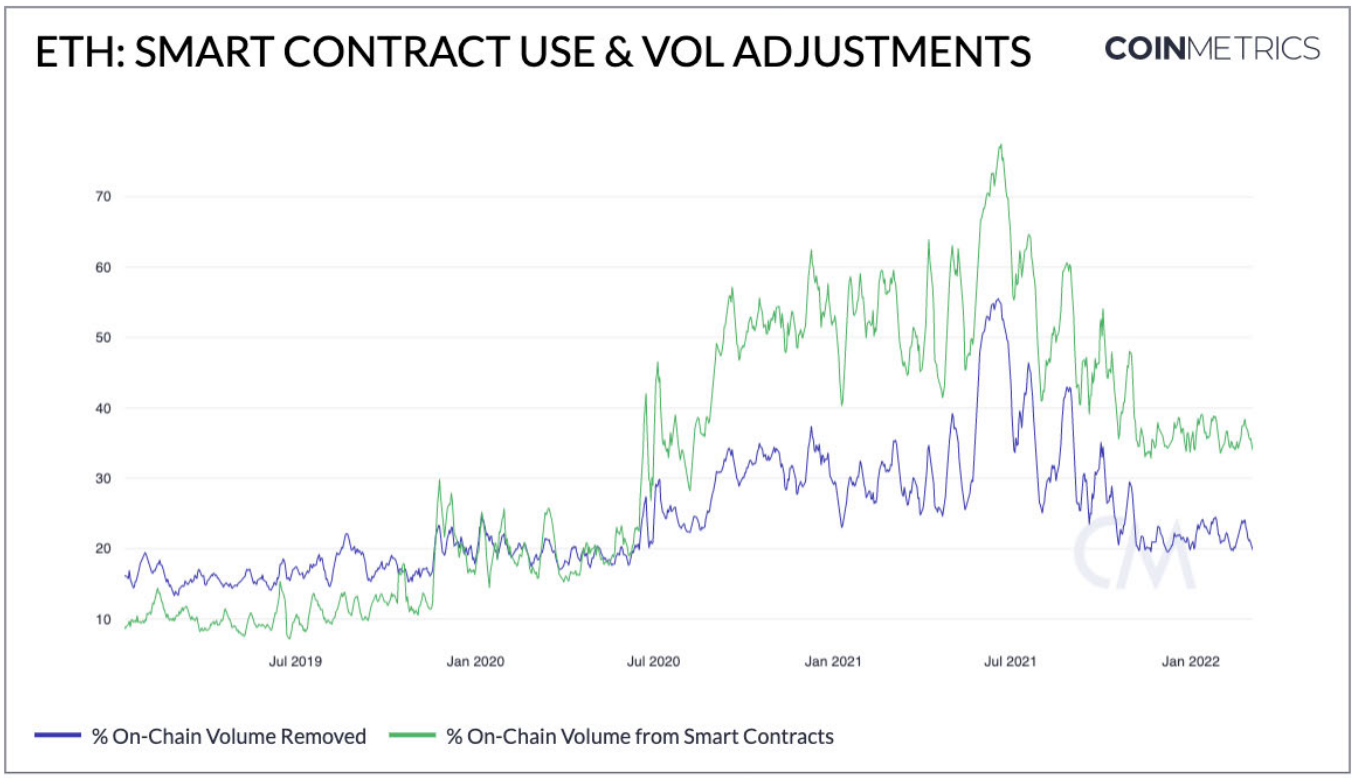
	Type	Baseline	Adjusted	% Removed
Bitcoin	UTXO	$13.1T	$4.7T	64%
Ethereum	Account	$5.4T	$3.6T	32%
Tether	Account (ERC20)	$2.5T	$1.9T	21%
USDC	Account (ERC20)	$2.1T	$1.4T	33%
DAI	Account (ERC20)	$0.8T	$0.6T	28%
Bitcoin Cash	UTXO	$2.8T	$0.5T	82%
Chainlink	Account (ERC20)	$153B	$113B	26%
Litecoin	UTXO	$0.6T	$0.1T	84%
Uniswap	Account (ERC20)	$86B	$66B	24%
Ethereum Classic	Account	$43B	$43B	0.4%

Source: Coin Metrics Network Data Pro

One observation is that the effect of heuristics is more evident with UTXO-based blockchains. As described earlier, this is likely due to the importance of accounting for change outputs.

Another interesting observation is the impact of smart contracts. On Ethereum, smart contracts are a special kind of account that contains code and data, and can also hold, send, and receive funds. Smart contracts often serve as autonomous intermediaries between parties when their functions are executed. Many decentralized finance (DeFi) transactions fit this criteria, an example being asset swaps on the automated market maker Uniswap. Ceteris paribus, we should expect smart contract activity to induce more pass through activity.

The chart below tells us that as relatively more ETH is transferred by smart contracts, more adjustments to total transfer volume are made. As DeFi activity accelerated during summer 2020, our pass through heuristic started filtering out more ETH transfer volume.



Source: *Coin Metrics' Formula Builder*

This observation is further reinforced by the small adjustments made to ETC volume on Ethereum Classic, created from the 2016 Ethereum hard fork. On any given day, a negligible amount of ETC is transferred via smart contracts, which aligns with our small adjustments made to total on-chain ETC volume.

Generally, we recommend that any cross-chain analysis of adjusted transfer volume be highly caveated. Because different heuristics are applied to UTXO and account-based chains, it is not appropriate to assume an apples-to-apples comparison. Some blockchains also have their own idiosyncrasies to account for, such as Cardano, which uses a modified UTXO model.

Ultimately, one-off spikes in transfer volume, even if adjusted, usually warrant further review. Better commentary can be made by supplementing transfer volume data with information like exchange holdings or average transfer sizes, or more granular data such as balance updates by specific entities and accounts.[7]

[7] Data from Coin Metrics' ATLAS block explorer, for example, can help observers hone in on specific transactions and accounts on-chain.

CONCLUSION

Accurately gauging the economic throughput of public blockchains today is essential for good decision making and analysis. It is clear today that public blockchains are undeniably conduits for meaningful economic activity. There are now countless real-world examples, such as the critical role that crypto has played to facilitate donations measured in the tens of millions of dollars in support of Ukraine and its people.

Yet, blockchain data inevitably contains some degree of noise, in the form of self-churn, spam, mixers and other non-meaningful activity. Not unlike macroeconomists who have long sought reliable methods in assessing the health of national economies, blockchain data observers need simple ways to isolate the signal from the noise. The heuristics described in this report are simple yet powerful rules that accomplish this goal.

Coin Metrics will continue to explore new ways to improve our estimates. Any new heuristic we introduce, though, will optimize across the dimensions of simplicity, transparency, accuracy, and consistency.[8]

As more eyes turn to crypto, understanding the power, and the constraints, of blockchain data analysis is becoming increasingly important. Coin Metrics looks forward to continue providing better estimates to the economic significance of public blockchains, and transparent methodologies and resources to better understand them.

[8] We avoid heuristics that might be effective but will likely add a *trend* to the data. One such example is looking for change outputs sent to different address types. We also avoid heuristics that we deem to have a high margin of error. We notably do not currently use any clustering algorithms to tag entity addresses as we believe it can introduce subjectivity to the data.

ON-CHAIN VOLUME

A DEEP DIVE INTO COIN METRICS' TRANSFER VALUE HEURISTICS

By **Nate Maddrey**, **Kyle Waters** and the **Coin Metrics Team**

COINMETRICS

To view more from Coin Metrics Research go to
coinmetrics.io/pubs or Subscribe to State of the Network