EFMA 2024 Plenary Panel on “Financial Big Data and Technology”

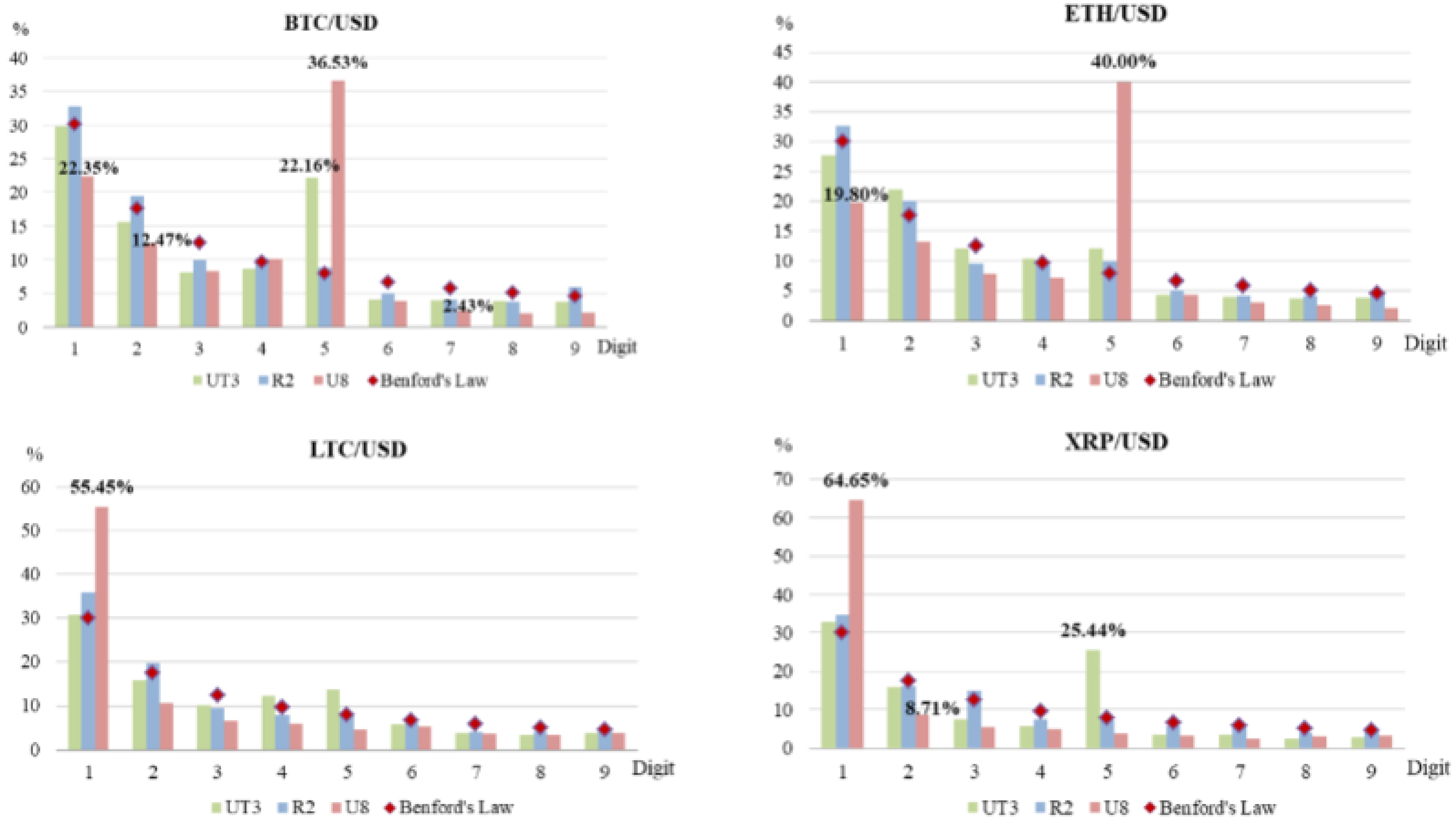
Web3 Big Data & Inclusion and Democratization Through Web3 and DeFi? Initial Evidence from the Ethereum Ecosystem

Lin William Cong
@Cornell University SC Johnson College of Business, IC3, and NBER
Ke Tang
@Institute of Economics, School of Social Science, Tsinghua University
Yanxin Wang & Xi Zhao
@School of Management, Xi'an Jiaotong University

Crypto Wash Trading

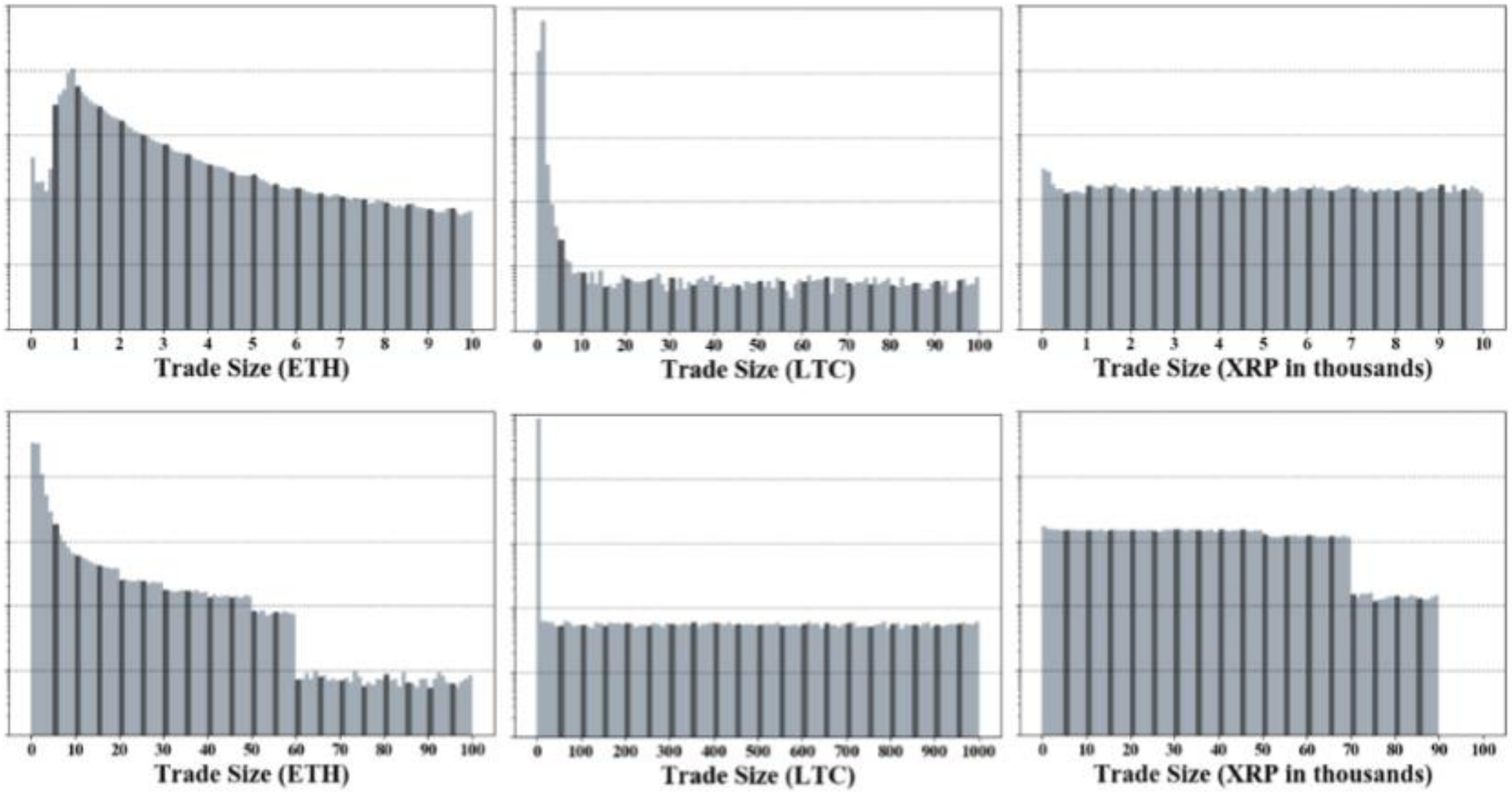
Traders fabricating trades and acting as the transaction counterparty on both sides:

Distribution of First Significant Digits



Rounding & Clustering

Unregulated tier-2 exchanges: U14



Quantifying Wash Trading

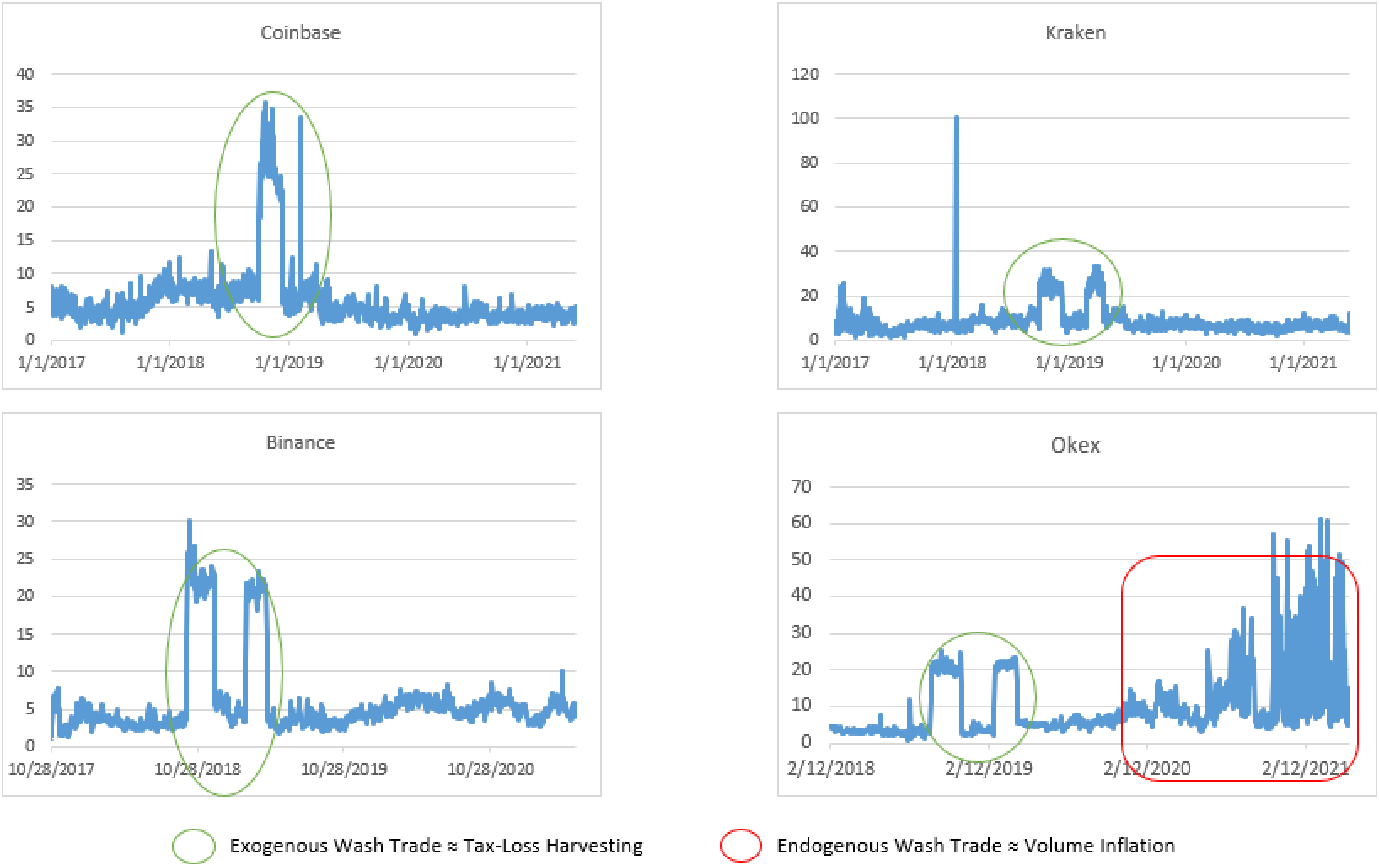
Round to unrounded trades ratio and regulated/traditional exchanges as benchmark.

	Wash Volume Percentage	
	Average	Standard Deviation
Unregulated exchanges	69.72%	29.71%
Unregulated Tier-1 exchanges	52.52%	29.41%
Unregulated Tier-2 exchanges	80.48%	25.13%

Exchange Code	Wash Volume Percentage
Panel A Unregulated Tier-1 Exchanges	
UT1	51.76%
UT2	51.73%
UT3	1.12%
UT4	92.60%
UT5	44.87%
UT6	66.3%
UT7	18.95%
UT8	66.12%
UT9	37.49%
UT10	94.31%

Panel B Unregulated Tier-2 Exchanges	
U1	99.99%
U2	98.30%
U3	72.72%
U4	95.50%
U5	89.71%
U6	98.13%
U7	77.20%
U8	77.09%
U9	81.12%
U10	98.45%
U11	21.48%
U12	98.08%
U13	65.42%
U14	96.78%
U15	94.36%
U16	23.27%

Tax-Loss Harvesting Evidence: BTC ↓



Estimated Tax-Loss Harvesting Revenue

Panel A - Tax-Loss Harvesting Estimates

Volume-Weighted		Equally-Weighted	
Harvest	**Regular**	**Harvest**	**Regular**
21.56	4.25	19.34	5.24

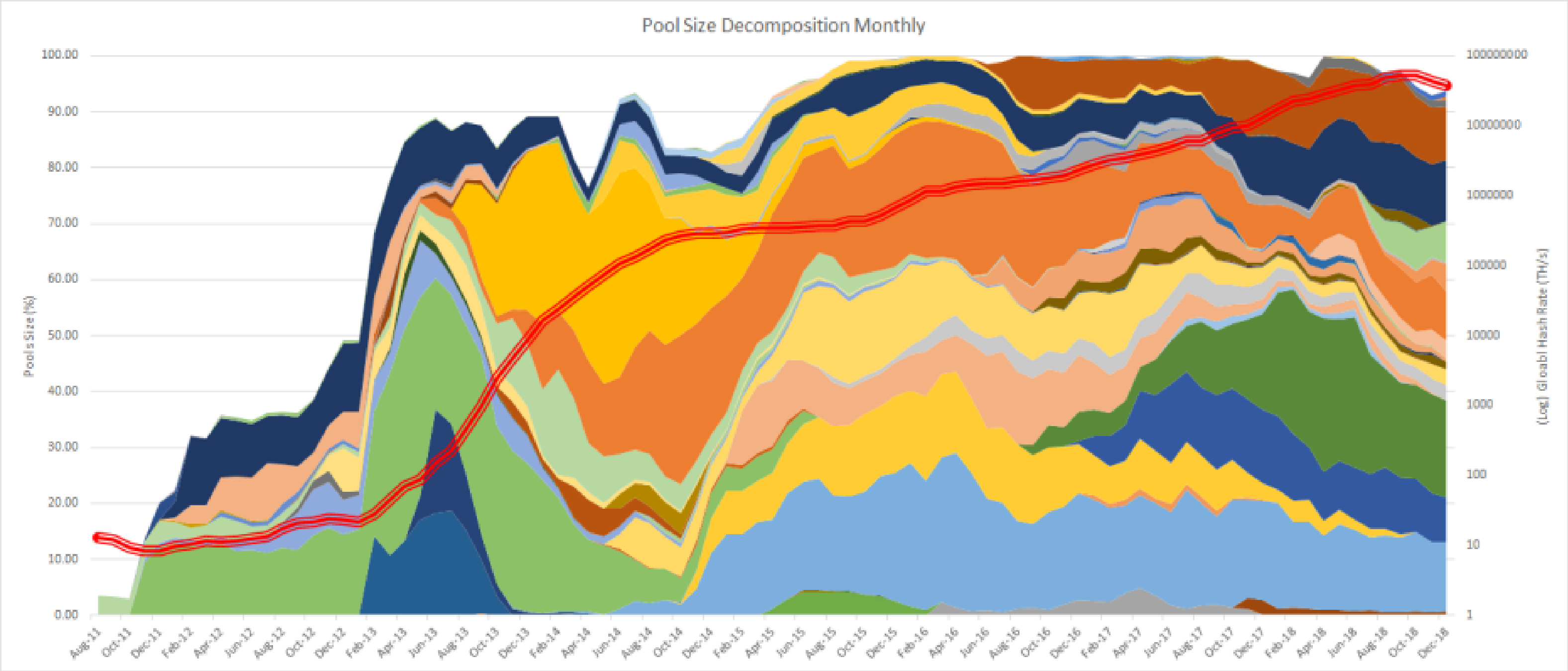
Panel B - Estimated Loss to the Government

Exchanges	Pair	Volume-Weighted		Equally-Weighted	
		Wash	**Revenue**	**Wash**	**Revenue**
All	BTC-USDT	25.52	5.36	20.80	4.37
Regulated	BTC-USDT	19.37	4.07	15.78	3.31
All	ALL	77.14	16.20	62.85	13.20
Regulated	ALL	58.53	12.29	47.69	10.02

Table 7. Estimating the Size of Revenue Loss from Tax-Loss Harvesting. Estimating the Size of Revenue Loss from Tax-Loss Harvesting. This table reports estimates of tax revenue loss arising from tax-loss harvesting in 2018. Panel A reports volume-weighted and equally-weighted estimates of the percentage of trades that are wash trades during tax-harvesting regular periods. Panel B reports the estimated wash volume and revenue loss to the government (in billions). All variables are reported at the regulated-exchanges level. See section 5.2 for computational details.

In 2018, federal capital gains tax revenue was **$158.4 billion**
-> Potential: **Increase** of about **5-10% tax revenue [only BTC]**.

Decentralized Mining in Centralized Pools (Cong, Li, and He, 2021)



1. Pool dominance coincides with explosive growth in hash power.
2. Pools grow but no long-term over-concentration.

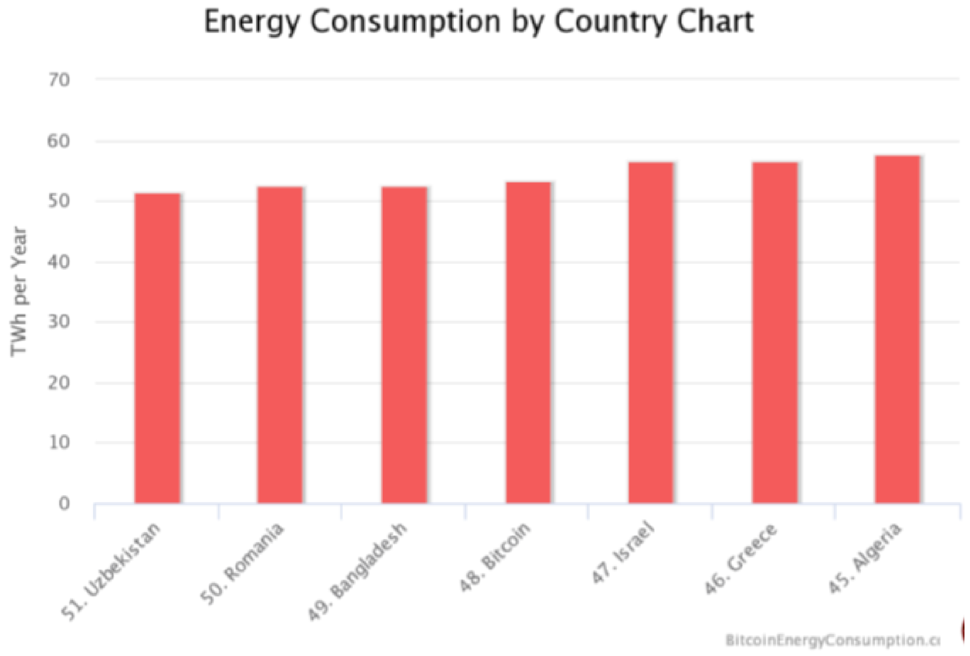
Decentralized Mining in Centralized Pools

- Risk-aversion → pooling: significant risk-sharing benefits.

1. Rise of pools not accompanied by over-concentration.
   - Diversification as a counter-centralization force.
   - IO force: larger pools charge higher fees and grow slower.
2. Financial innovation that potentially reduces welfare.
   - Risk-sharing drastically aggravates mining arms race and multiplies egregious energy use.

Vertical Integration and Mining Pools

1. Vertically integrated entities (e.g., exchanges such as FTX).
2. Liability run and asset spiral.
3. Mining Concentration and environmental damages:
   - Consensus protocol relies on adequate decentralization for security (e.g., 51% attack, selfish mining, etc.
   - Technological possibility or economic reality?
   - Pooled mining with enormous energy consumption.



An Anatomy of Crypto-Enabled Cybercrimes (Cong, Harvey, Rabetti, and Zong, 2022)

- *Aggregate users to learn group interactions (network analysis).*
- *Trace specific transactions (e.g., ransom payments);*
- *Learn the economics of a group of users (.e.g, cybercrime);*
- *Infer an economic activity from transaction's patterns (e.g., tax evasion);*
- *Combine with other data sets to provide perspectives on crypto adoption and usage (e.g., financial inclusion).*

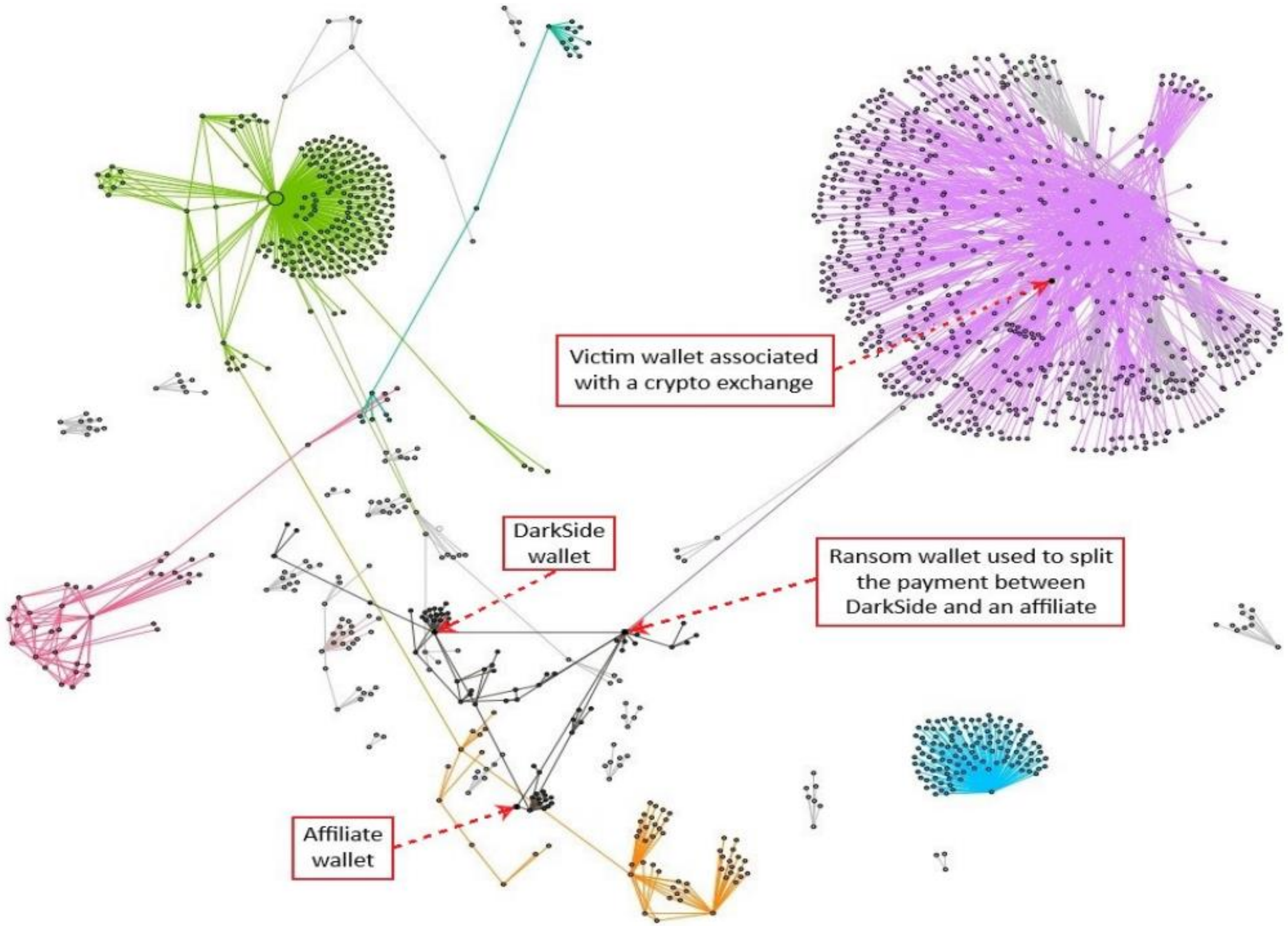


Figure 5: A Ransomware Gang's (DarkSide) Network Analysis

Ransomware

- *Ransomware refers to a cyberattack in which criminals use malware to encrypt all of the files on the victim's device or network, making them inaccessible;*
- *The attacker will then demand a payment, nowadays always in digital currencies, in exchange for a decryption key needed to regain access to the files; multiple layers of extortion*
- *Ransomware attacks can render victim organizations virtually inoperable, and attackers often target critical infrastructure organizations including banks, energy providers, hospitals, schools, and municipal governments.*
- *Organized crimes and underreporting.*

U.S. DEPARTMENT OF THE TREASURY

ABOUT TREASURY POLICY ISSUES DATA SERVICES NEWS

PRESS RELEASES

Treasury Sanctions Evil Corp, the Russia-Based Cybercriminal Group Behind Dridex Malware

December 5, 2019

Washington – Today the U.S. Treasury Department's Office of Foreign Assets Control (OFAC) took action against Evil Corp, the Russia-based cybercriminal organization responsible for the development and distribution of the Dridex malware. Evil Corp has used the Dridex malware to infect computers and harvest login credentials from hundreds of banks and financial institutions in over 40 countries, causing more than $100 million in theft. This malicious software has caused millions of dollars of damage to U.S. and international financial institutions and their customers. Concurrent with OFAC's

Rebranding Strategy

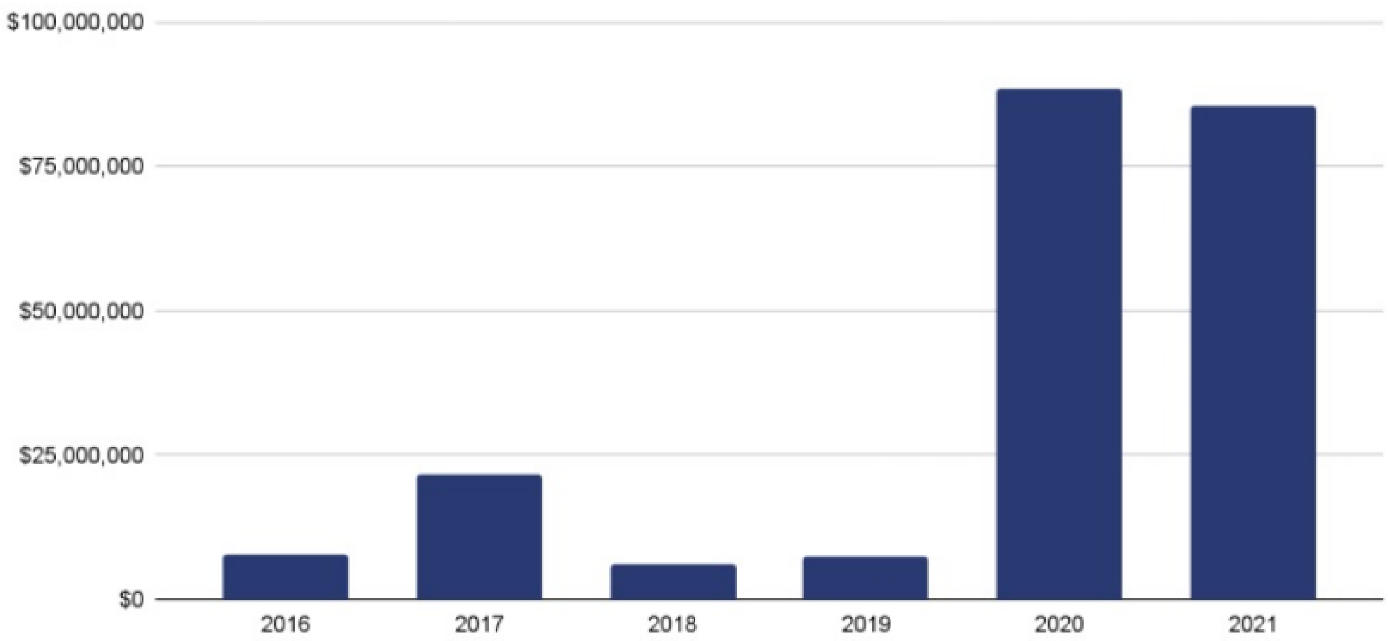


Fig. 1.8 Ransomware payment rule to strains associated with Evil Corp, 2016-2021.

Overview:

Financial inclusion and democratization through Web3 and DeFi?

"A Fundamental New Approach to corporate governance, value creation and stakeholder participation with pari passu interests." (WEF)

Functional
Efficient
Fair
Affordable

- Description of Ethereum Ecosystem using Big Data
  - Network Structure & General Trends/Stylized Pattern
  - Distributions of Mining Income, Onchain Wealth, and Utilization
  - Data sharing and visualization portal

- Transaction Fees and Financial Inclusion/Democratization
  - Percentage Transaction Fee; Network Congestion and Gas Price; Fee and Extra Gas Fee Reserved; Transaction Failures; Token Exchange Rate Volatility

- Inclusion and Democratization Through Redistributive "Monetary" Policy (fee mechanism changes and programs e.g. ,airdrops).

Literature

- **Transaction fees in distributed networks:**
  - Analyzing transaction fee and relating it to congestion and system stability**:** Easley, O'Hara and Basu (2019) and Huberman, Leshno and Moallemi (2021).
  - Transaction fee design: SPA, Basu et al. (2019), **EIP-1559** (Roughgarden, 2020b; Reijsbergen et al., 2021; Liu et al., 2022).
  - Fees on Dex: John, Rivera, and Saleh (2022); Capponi, Jia, and Yu (2022).

- **DeFi and Web3:**
  - **Future of finance?** Harvey, Ramachandran and Santoro (2020).
  - **DeFi applications** such as Decentralized Exchanges and automated market-making (Lehar and Parlour, 2021; Capponi and Jia, 2021; Park, 2021; Augustin, Chen-Zhang and Shin, 2022) or lending (Markovich, 2021; Lehar, Parlour and Berkeley, 2022).
  - **Blockchain ecosystem** (e.g., Cong, He and Li, 2018; Rosu and Saleh, 2021; Makarov and Schoar, 2022; Zhang, Ma, and Liu, 2022).

- **Digitization, transaction costs, and financial inclusion:**
  - Philippon (2016); Jack and Suri (2014); Bachas et al. (2018).

Big Data and Big Computation

- Ethereum blockchain
  - Aug 15-Feb 22; 14 million blocks, 1.7+4.6 billion transactions, 1 billion transfers, 433 DeFi, 5047 Dapps, etc.;
  - Value of tokens transferred, the time when transaction bundled into the block, gas used, gas price and gas limit (set by the initiator), status of transaction.
- Block information (e.g., address of block verifier, mining pool, block number, etc.)
- Addresses associated with DeFi/ DApps
  - DApp Radar, DApponline, and Etherscan
  - Classified into 9 groups: exchanges, DeFi, gambling, games, collectibles, etc.
  - 166 DeFi protocols, 2,820 DApps.
- ETH Gas Station, CoinMarketCap, Google Trends
  - Recommended gas prices, etc.
  - Token prices, popularity metrics, etc.
- Large-scale computation: 14 servers dual Xeon E5, 128G Mem, 48TB hard disks,

1. The Ethereum Ecosystem

Ethereum Ecosystem (exchanges and DeFi dominates)

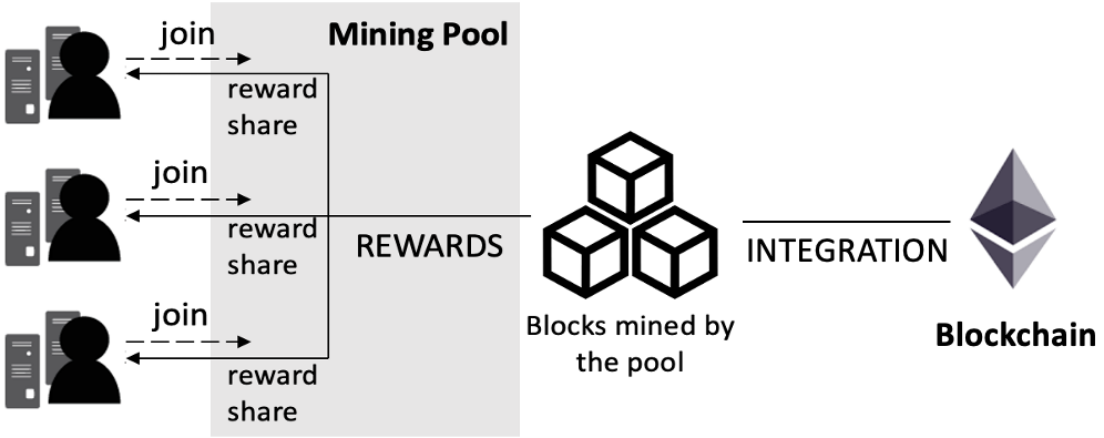
Two types of accounts

Externally Owned Account (EOA)

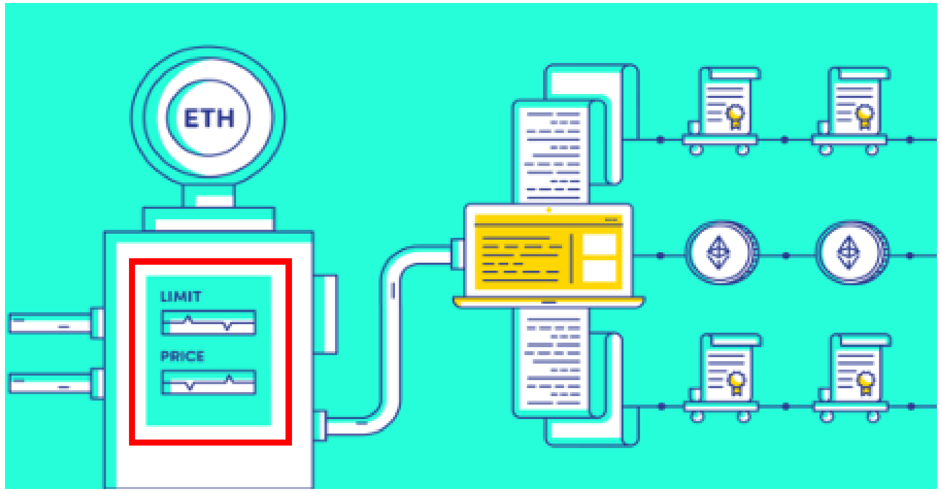
Controlled by a private key

Contract Account (CA)

Smart contract with automatically executed codes



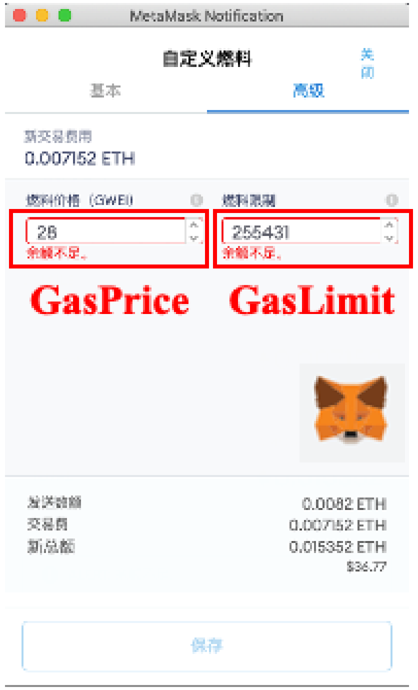
Gas Mechanism



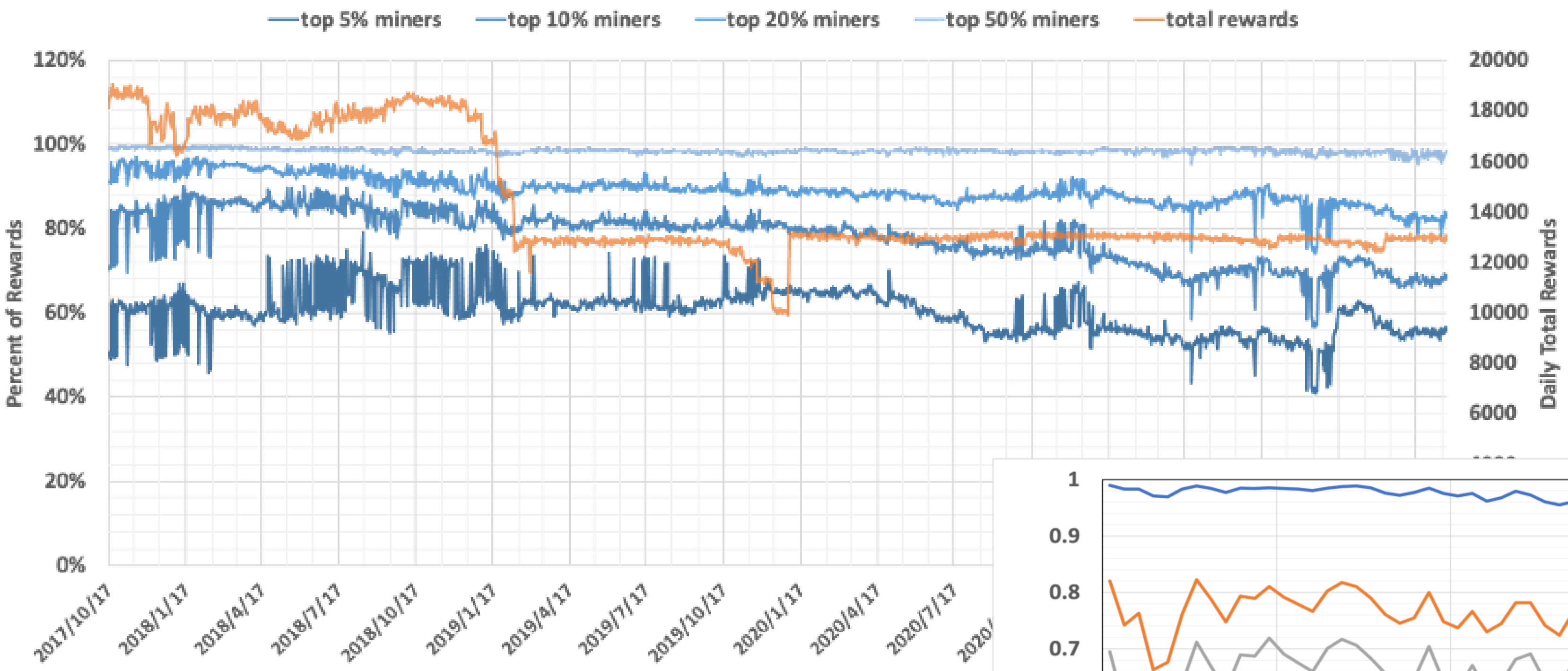
Gas Related Parameters

- Gas limit } User set
- Gas price } User set
- Gas used
- Transaction fee = GasPrice * GasUsed

*Note: one ought to reserve enough ETH in his/her wallets (GasPrice * GasLimit)*



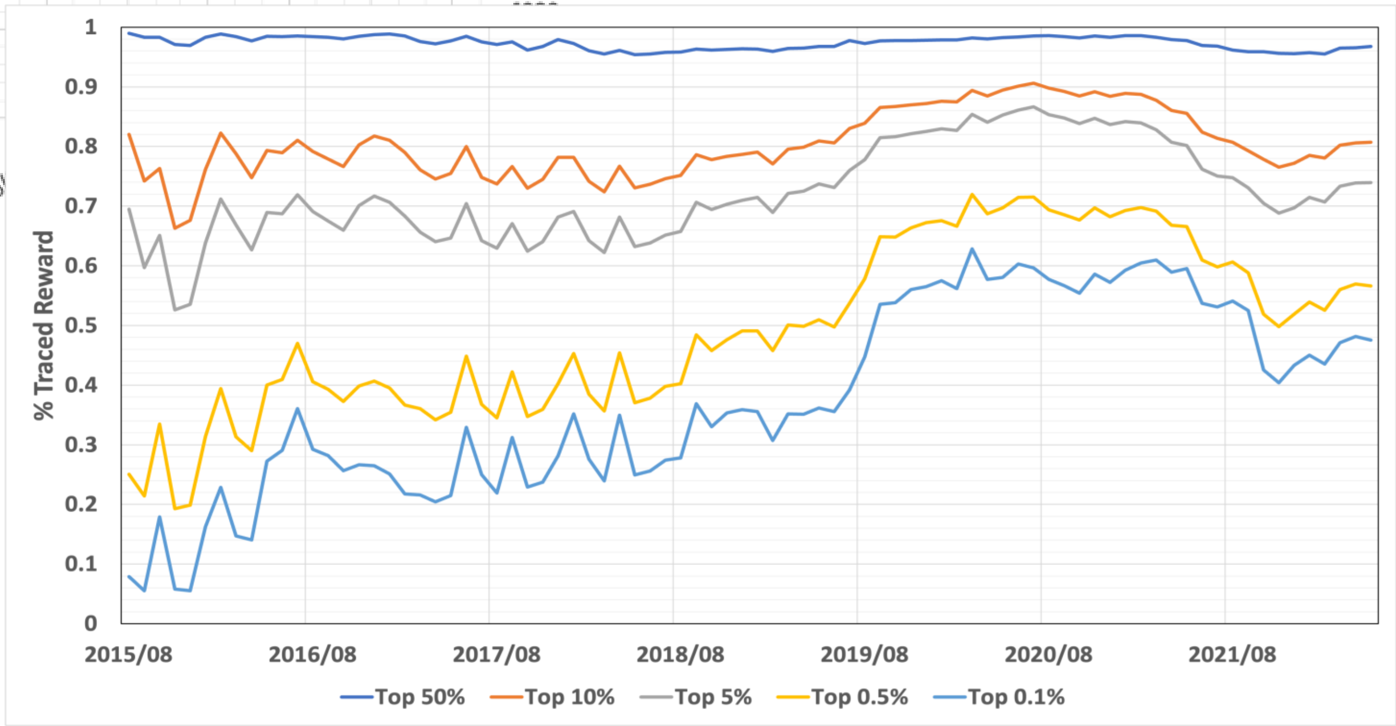
Description of Ethereum Ecosystem—*Distribution of Mining Income*



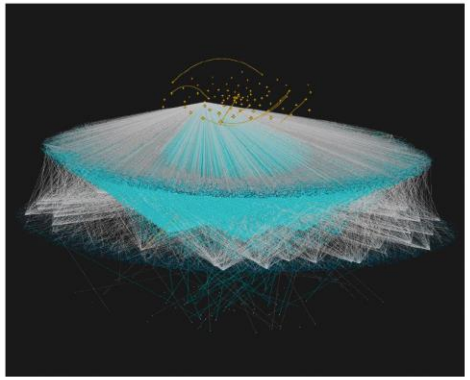
Mining Rewards Received by Mining Pools

- Mining activities and rewards are concentrated in large mining pools and nodes, similar to Bitcoin.

Traced Mining Rewards for Miners



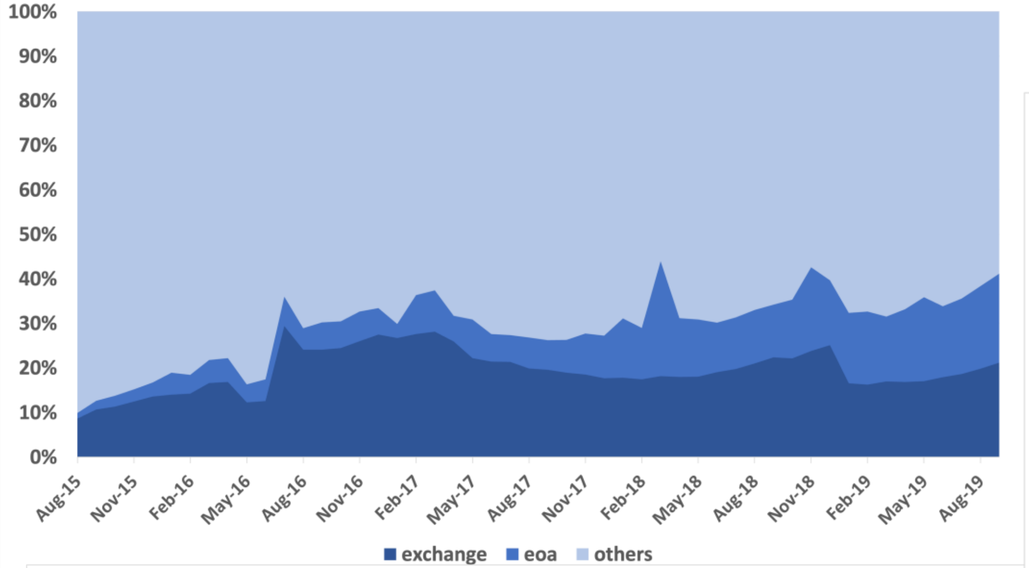
The Tracing Porcess

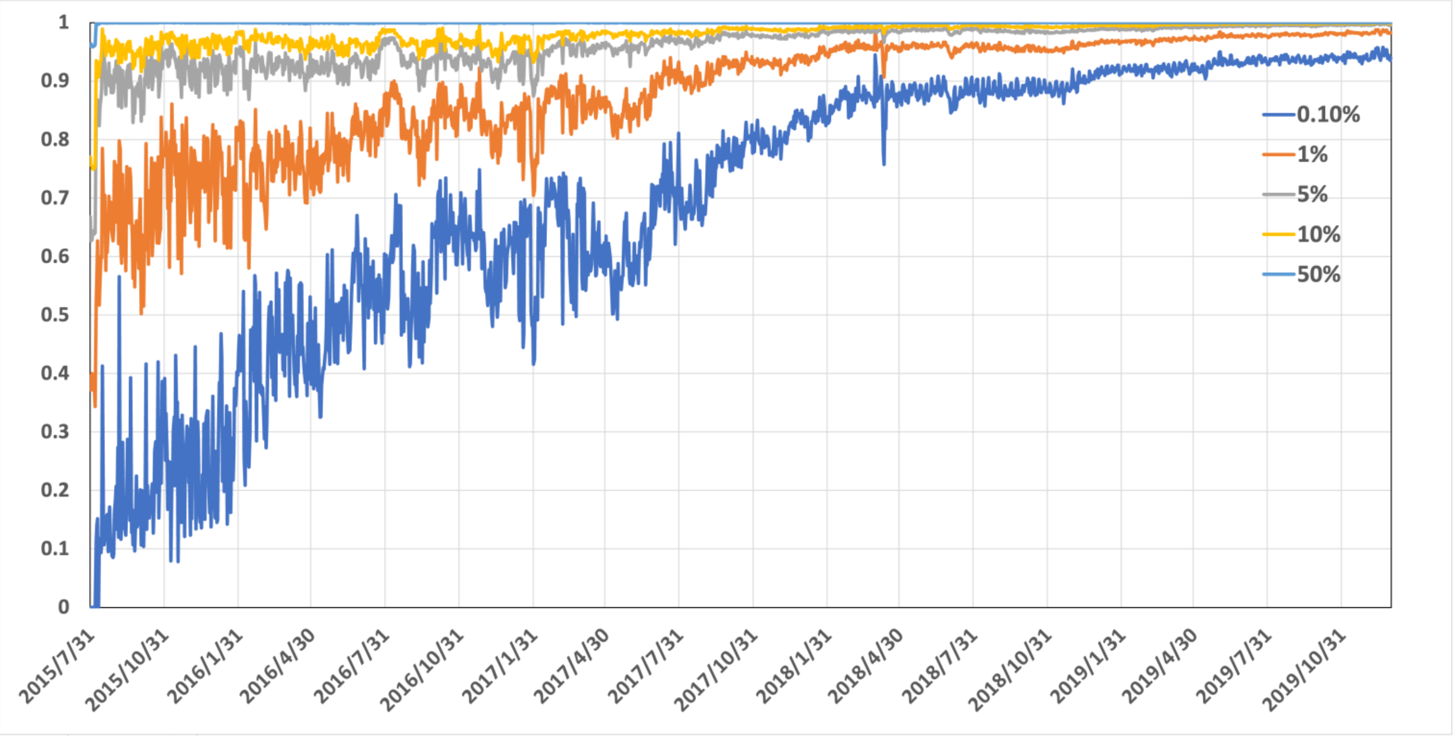


Description of Ethereum Ecosystem—*Distribution of On-Chain Token Ownership*

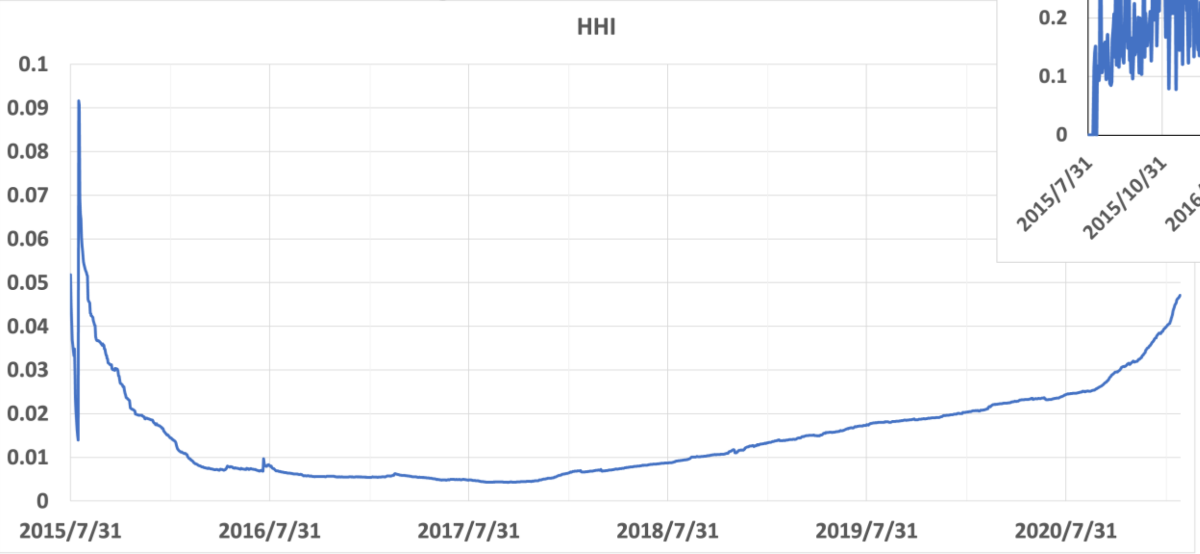
- Token ownership is heavily concentrated at a few nodes of institutions and individual users.

Distribution of Ether holding among different nodes





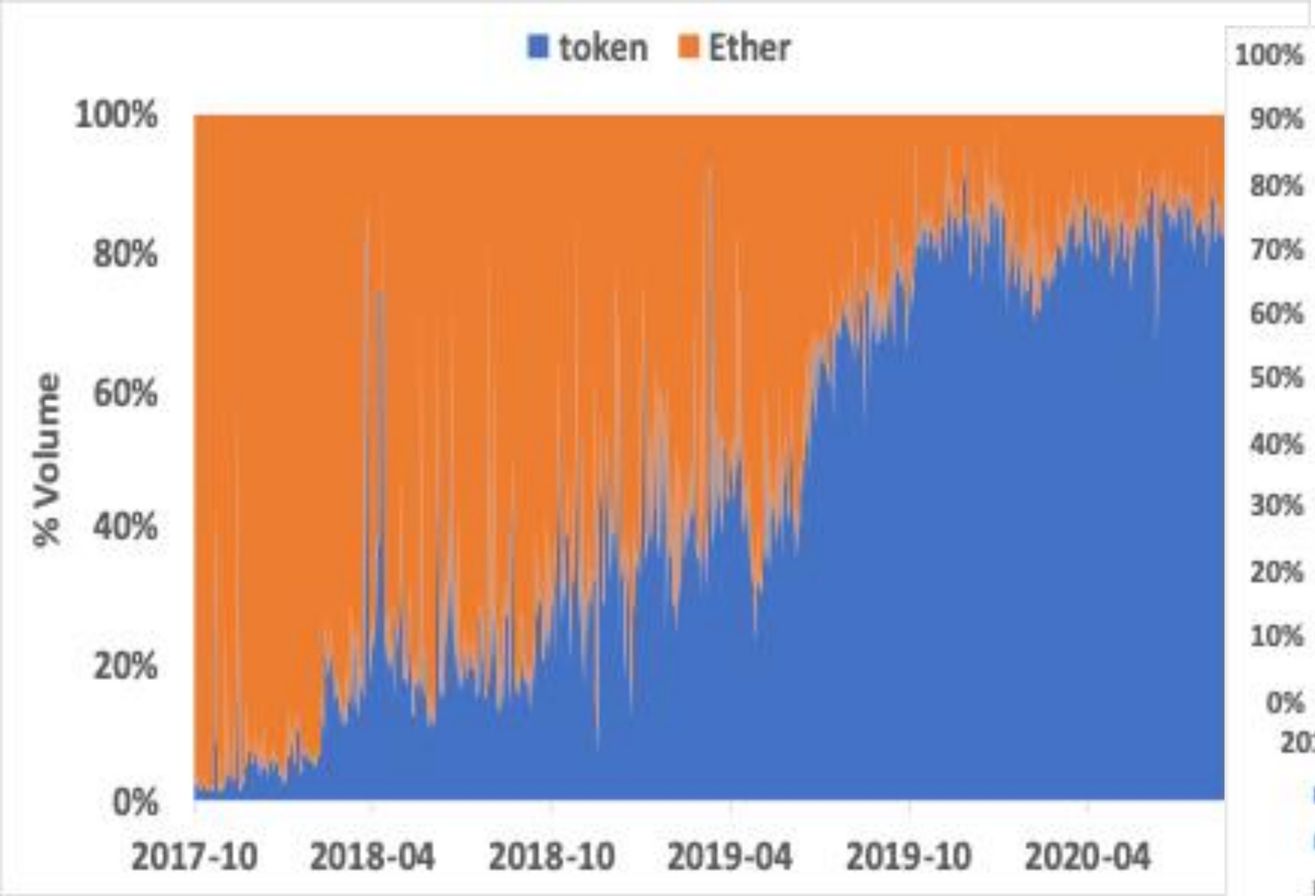
The concentration of EOA addresses (without exchanges)

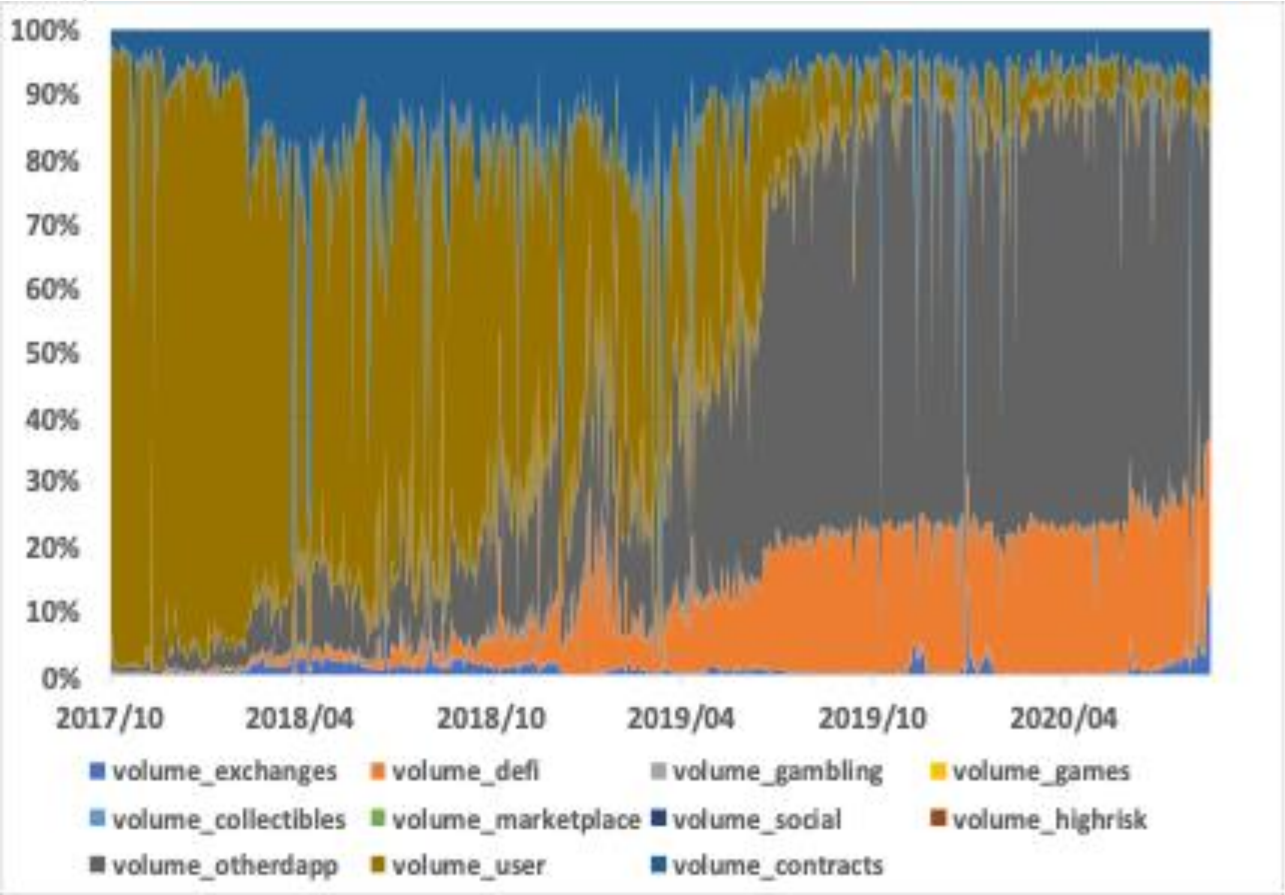


HHI for EOA address (without exchanges)

Description of Ethereum Ecosystem—*Distribution of Transactions*

- Shift from peer-to-peer interactions to user interactions with Dapps and DeFi protocols.
- Significantly more transactions by large players.





Decomposition of Transaction Volume

2. Inclusion and Democratization?
A Transaction Fee Perspective

Fundamentally about technology and fee design, not IO and market power.

Transaction Fees and Undemocratic and Exclusive Usage—*Percentage Transaction Fee*

$$PercentageTransactionFee = \frac{GasPrice * GasUsed}{Value} \times 100\%$$

- The percentage transaction fee for small amount transactions using DeFi is too high and volatile for inclusive finance.

Table 2—: Percentage Transaction Fee (continued)

(c) Ether and Tokens on Ethereum

	Percentage transaction fee of transactions with Ether						Percentage transaction fee of transactions with tokens					
value ($)	mean (%)	median (%)	25% (%)	75% (%)	standard deviation	count	mean (%)	median (%)	25% (%)	75% (%)	standard deviation	count
0-0.01	2.05*1016	1549.53	121.75	6.4*104	3.38*1015	1,802,606	6.56*1031	15757.34	2108.71	8.68*105	3.16*1032	1,020,664
0.01-0.1	150.45	37.82	21.00	70.00	39.14	10,828,833	863.32	239.92	87.87	384.17	37.96	3,096,112
0.1-1	31.54	16.80	7.19	32.38	6.49	33,110,009	96.47	29.41	9.86	76.24	8.43	5,838,297
0-1	8.07*1014	21.00	10.11	44.10	6.73*1014	45,741,448	6.68*1030	69.84	18.11	287.03	1.01*1032	9,955,073
1-10	7.81	2.11	0.42	8.75	7.60	53,548,484	17.88	4.15	1.42	11.45	2.54	10,608,388
10-100	1.24	0.15	0.04	0.64	2.01	109,237,500	2.53	0.58	0.21	1.67	0.19	23,077,554
100-1000	0.18	0.04	0.01	0.13	2.19	78,726,642	0.36	0.09	0.03	0.26	0.01	43,924,023
1000-	0.02	0.00	0.00	0.01	0.03	52,759,079	0.05	0.01	0.00	0.03	0.00	38,500,612
1-	1.93	0.08	0.02	0.53	3.65	294,271,705	2.29	0.08	0.01	0.51	0.78	116,110,577
General	1.09*1014	0.13	0.02	1.84	2.47*1014	340,013,153	5.29*1029	0.11	0.02	0.91	2.84*1031	126,065,650

Linking Transaction Fee to Network Utilization, Gas Fees, and Extra Gas Fee Reserved

- Significant negative relationship between gas price and delay time.
- Increases in gas price, transaction value, and network congestion all predicts increases extra gas reserve.
- 1% increase of network utilization predicts in an additional 3.43% gas price for all transactions

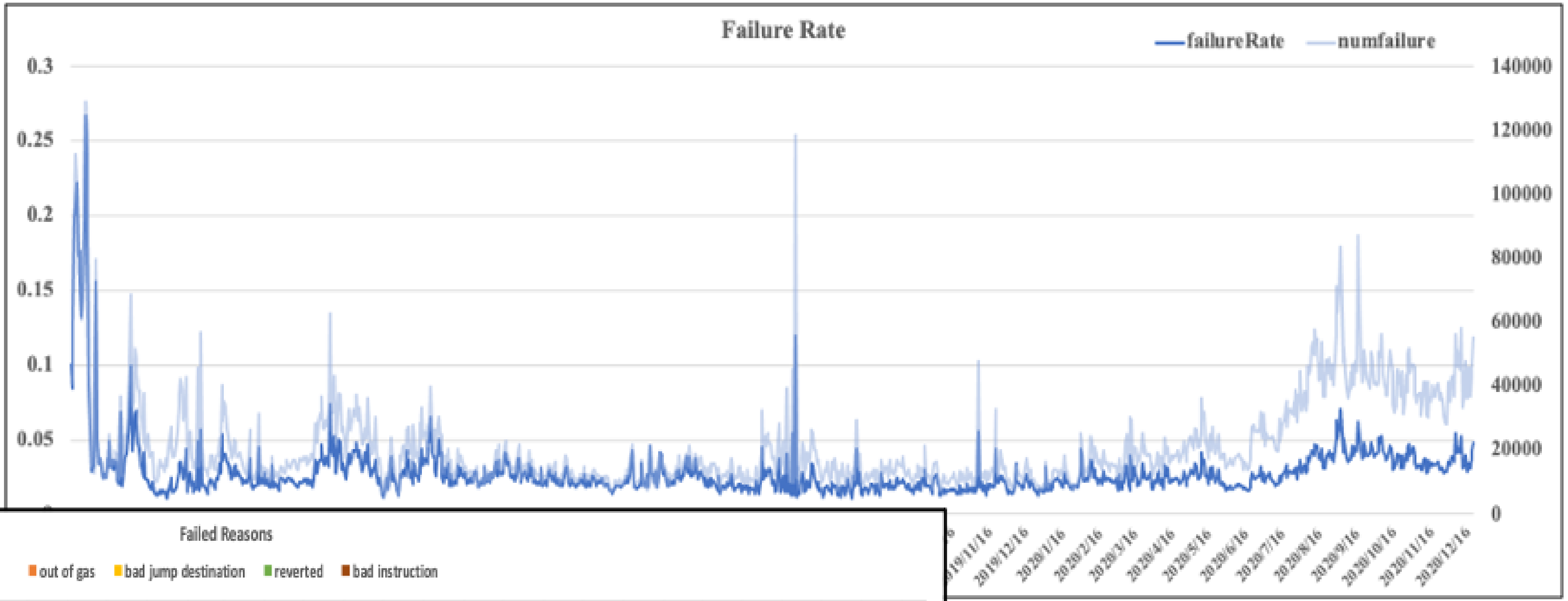
Determinants of Extra Gas Reserved

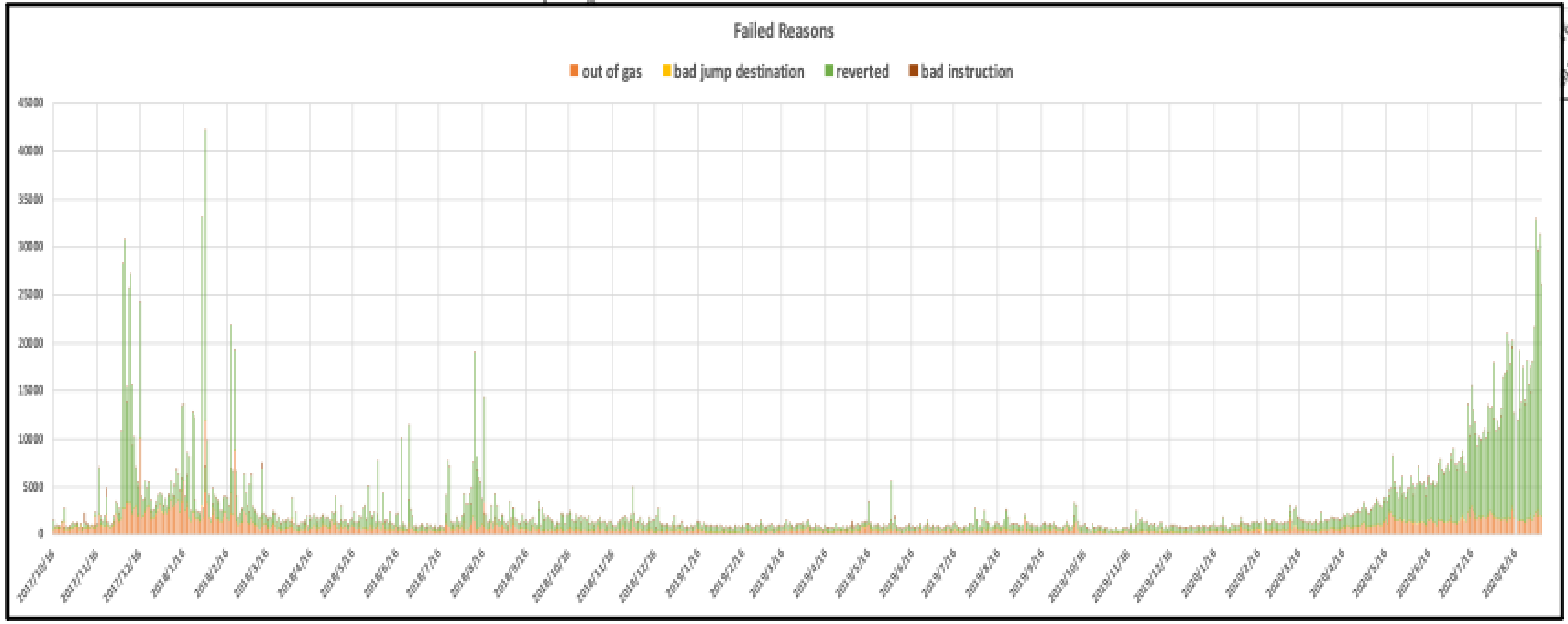
Ln(ExtraGasReserved)	All
L.Ln(NetworkUtilization)	0.409***
	0.002
L.EthReturn	-0.695***
	0.003
L.ln(MedianGasPrice)	0.048***
	0.000
L.ln(BlockRewards)	-0.990***
	0.001
L.ln(EthPopularity)	-0.076***
	0.000
Obs.	753,191,813
AIC	$4.622*10^{9}$
Null Deviance	$2.043*10^{10}$

(a) How Much Users Need to Reserve in the Wallets

	mean	median	25%	75%	standard deviation	
ExtraGasFee ($)	5.455	0.077	0.00	1.559	37.049	753,191,813
GasFee ($)	4.075	0.434	0.068	2.701	135.535	753,191,813

Transaction Failures





- **Major Reasons for Failure**
  a) "Out of gas"(30%)
  b) "Reverted" (73%)
  c) "Bad Instruction"
  d) "Bad jump destination

3. Inclusion/Democratization: EIP 1559 and Airdropping as Monetary Redistribution Policies

The EIP-1559 Fee Mechanism—*Background and Identification Strategy*

- **EIP-1559**
  - ✓ Burnt base fee
  - ✓ Max priority fee and priority fee (tips)
  - ✓ Block size

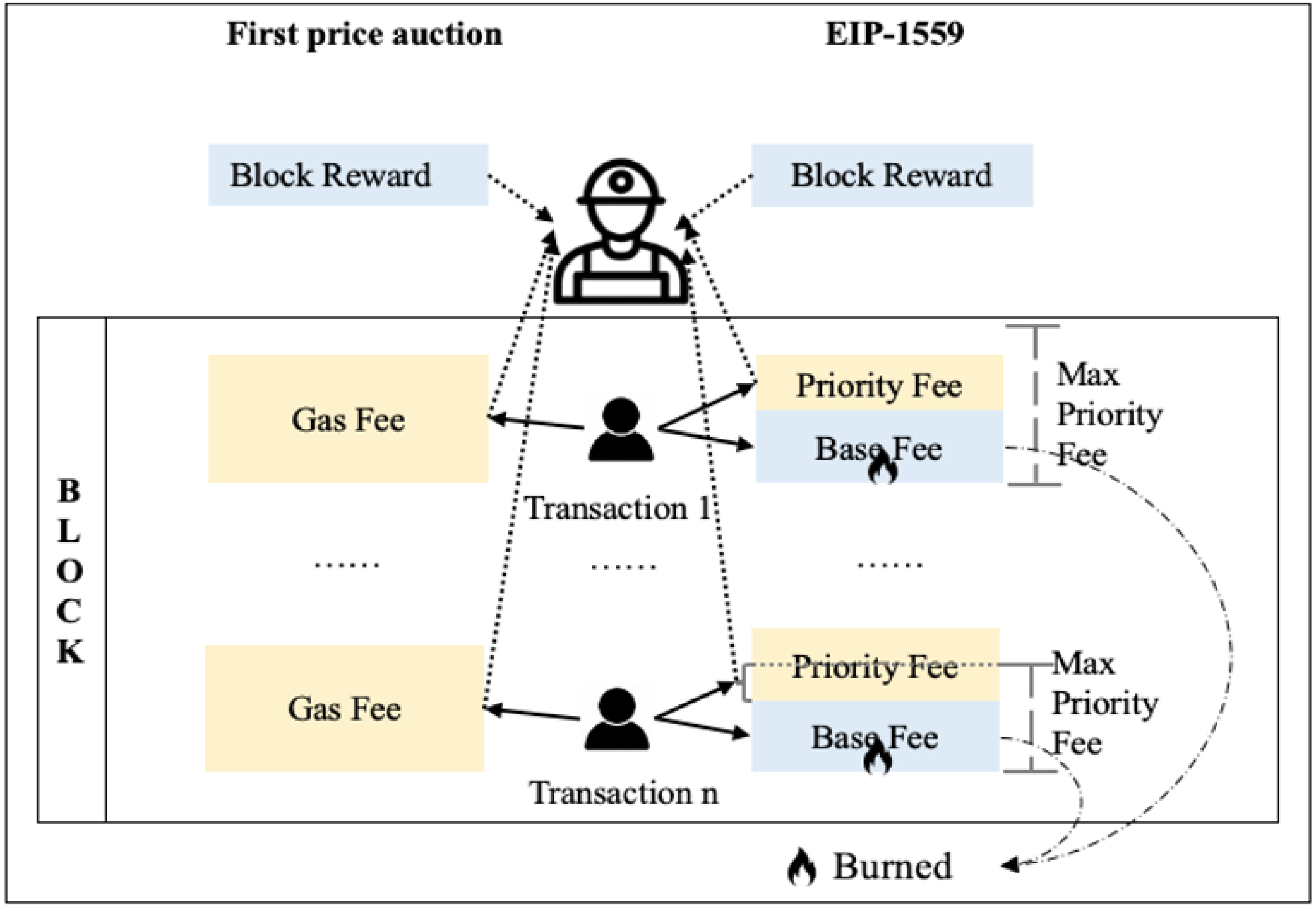
- **Identification Strategy**
  - ✓ **Regression discontinuity in time** (RDiT) to estimate overall effects.

$$y_{it} = \alpha + \beta Burning_{it} + \gamma f(date_{it}) + \delta X_{it} + \varepsilon_{it}$$

  - ✓ Generalized **difference-in-difference** method to account for the heterogeneity of miners and users.

$$y_{mt} = \beta Ln(PercentBlock_m) \times Burning_t + \omega X_{mt} + \lambda_m + \gamma_t + \varepsilon_{mt}$$

EIP-1559 Fee Mechanism



Redistributive Effect of EIP-1559—*Miner Side*

- **Overall negative effect** on miners' mining rewards.
- Weekly rewards for miners belong to **larger mining pools decreased less.**
- Weekly rewards for miners with **higher computation power decreased more.**

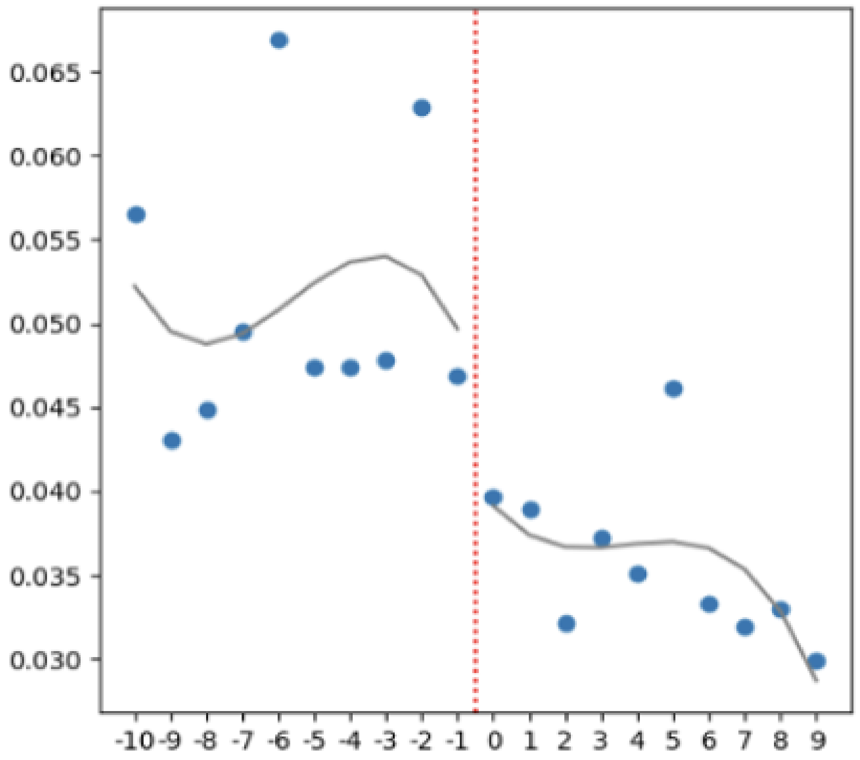
Overall Effects of EIP1559 on Mining Rewards

	Main		Exclude a week		Exclude two weeks	
	(1)	(2)	(3)	(4)	(5)	(6)
LnRewards	10 weeks	20 weeks	10 weeks	20 weeks	10 weeks	20 weeks
Burning	-0.007***	-0.008***	-0.007***	-0.006***	-0.008***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Observations	2,709,380	5,418,760	2,438,442	5,147,822	2,167,504	4,876,884
R-squared	0.020	0.058	0.022	0.060	0.019	0.062
Number of miners	135,469	135,469	135,469	135,469	135,469	135,469
Controls	YES	YES	YES	YES	YES	YES
Miners FE	YES	YES	YES	YES	YES	YES
Month FE	NO	NO	NO	NO	NO	NO

Robust standard errors in parentheses

*** $p<0.001$, ** $p<0.01$, * $p<0.05$

Redistributive Effect of EIP-1559—*Miner Side*



(a) The Log of Weekly Mining Rewards

VARIABLES	(1) 20 weeks	(2) 20 weeks	(3) 10 weeks	(4) 10 weeks
LnPercentBlocks*Burning	0.056***		0.010***	
	(0.004)		(0.002)	
LnBeforeRewards*Burning		-0.068***		-0.029***
		(0.001)		(0.001)
LnMiners	0.009***	0.008***	0.017***	0.016***
	(0.001)	(0.001)	(0.001)	(0.001)
LnGasprice	0.013***	0.013***	0.008***	0.008***
	(0.000)	(0.000)	(0.000)	(0.000)
LnDeviantGasprice	0.001***	0.001***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
LnEtherprice	0.005***	0.004***	-0.009***	-0.011***
	(0.001)	(0.001)	(0.001)	(0.001)
LnDifficulty	-0.091***	-0.093***	-0.028***	-0.030***
	(0.001)	(0.001)	(0.002)	(0.002)
LnCongestion	0.015***	0.017***	0.085***	0.090***
	(0.002)	(0.002)	(0.004)	(0.004)
Observations	5,418,760	5,418,760	2,709,380	2,709,380
R-squared	0.080	0.185	0.030	0.065
Number of miners	135,469	135,469	135,469	135,469
Miners FE	YES	YES	YES	YES
Month FE	YES	YES	YES	YES

Robust standard errors in parentheses
*** $p<0.001$, ** $p<0.01$, * $p<0.05$

Heterogenous Effects of EIP1559 on Mining Rewards

Redistributive Effect of EIP-1559—*User Side*

- Overall **positive effect** on transaction volume and no. of Dapps used.
- Significant negative coefficients of interaction terms:
  Users with **lower transaction frequency** or **less ETH balance benefit more.**

(a) Weekly Transaction Volume

	Main		Exclude a week		Exclude two weeks	
	(1)	(2)	(3)	(4)	(5)	(6)
LnVolume	10 weeks	20 weeks	10 weeks	20 weeks	10 weeks	20 weeks
Burning	0.002***	0.003***	0.003***	0.004***	0.005***	0.006***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)
Observations	5,045,800	10,091,600	4,541,220	9,587,020	4,036,640	9,082,440
R-squared	0.000	0.002	0.000	0.002	0.000	0.002
Number of users	252,290	252,290	252,290	252,290	252,290	252,290
Controls	YES	YES	YES	YES	YES	YES
Miners FE	YES	YES	YES	YES	YES	YES
Month FE	NO	NO	NO	NO	NO	NO

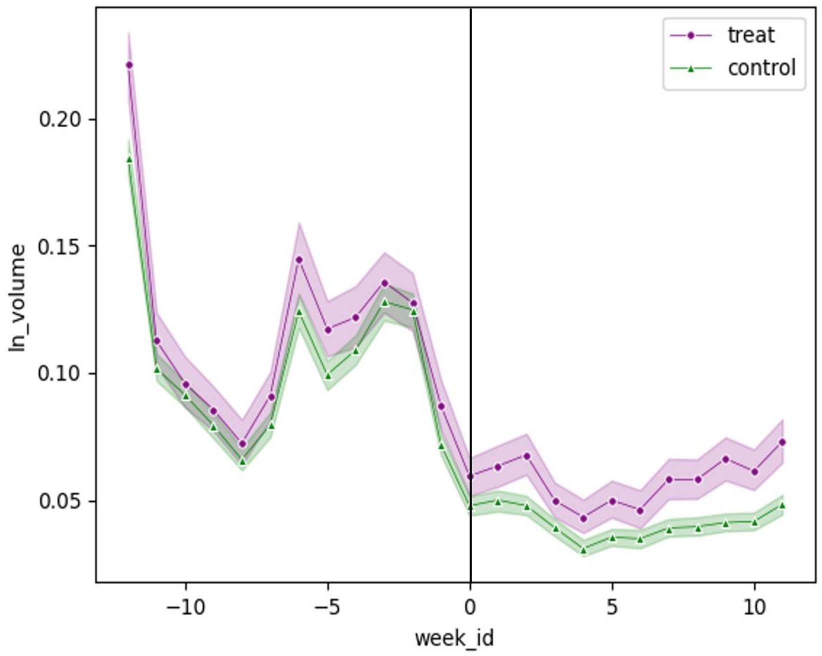
Robust standard errors in parentheses
*** $p<0.001$, ** $p<0.01$, * $p<0.05$

Inclusion and Democracy Through Airdropping (+ Impact on ETH Price)

- **Background**
  - OmesiGo: First large-scale airdrop on Ethereum, Sept 13-23, 2017.
  - Airdropping OMG to addresses with Ether balance > 0.1 ETH at block height 3988888.

- **Identification Strategy:**
  - DiD with RD sample.
  - SCM (synthetic ETH).

$$y_{it} = \beta(After_{it} \times Airdrop_i) + \omega X_{it} + \lambda_i + \gamma_t + \varepsilon_{it}$$



The Impact of Airdrop on Users' Weekly Transaction Volume

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	bandwidth 0.015	bandwidth 0.015	bandwidth 0.01	bandwidth 0.01	bandwidth 0.005	bandwidth 0.005
after_airdrop	0.038***	0.035***	0.037***	0.033***	0.038***	0.034***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
after	-0.101***		-0.102***		-0.097***	
	(0.002)		(0.002)		(0.002)	
Observations	880,771	880,771	760,608	760,608	585,100	585,100
R-squared	0.010	0.013	0.011	0.013	0.011	0.013
Number of miner_id	36,700	36,700	31,693	31,693	24,380	24,380
Controls	NO	YES	NO	YES	NO	YES
Weighted	YES	YES	YES	YES	YES	YES
Miners FE	YES	YES	YES	YES	YES	YES
Month FE	NO	YES	NO	YES	NO	YES

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Conclusions

- Web3 and DeFi widely advocated as innovations for greater inclusion and democratization.
- First comprehensive description (utilizing big data) of the Ethereum ecosystem including its network structure, trends, and distributions of mining, ownership, and transactions.
- Fee mechanisms are not conducive to inclusion and democratization due to discrimination against small/pool players, high failure rate, etc.
- Protocols changes and programs such as base fee burning (EIP 1559) and airdropping represents redistributive "monetary" policies and can improve financial inclusion and democratization.
- Source of information and initial benchmark; future research needed.