Decentralized Mining in Centralized Pools

Will Cong — Cornell
Zhiguo He — Chicago Booth & NBER
Jiasun Li — George Mason

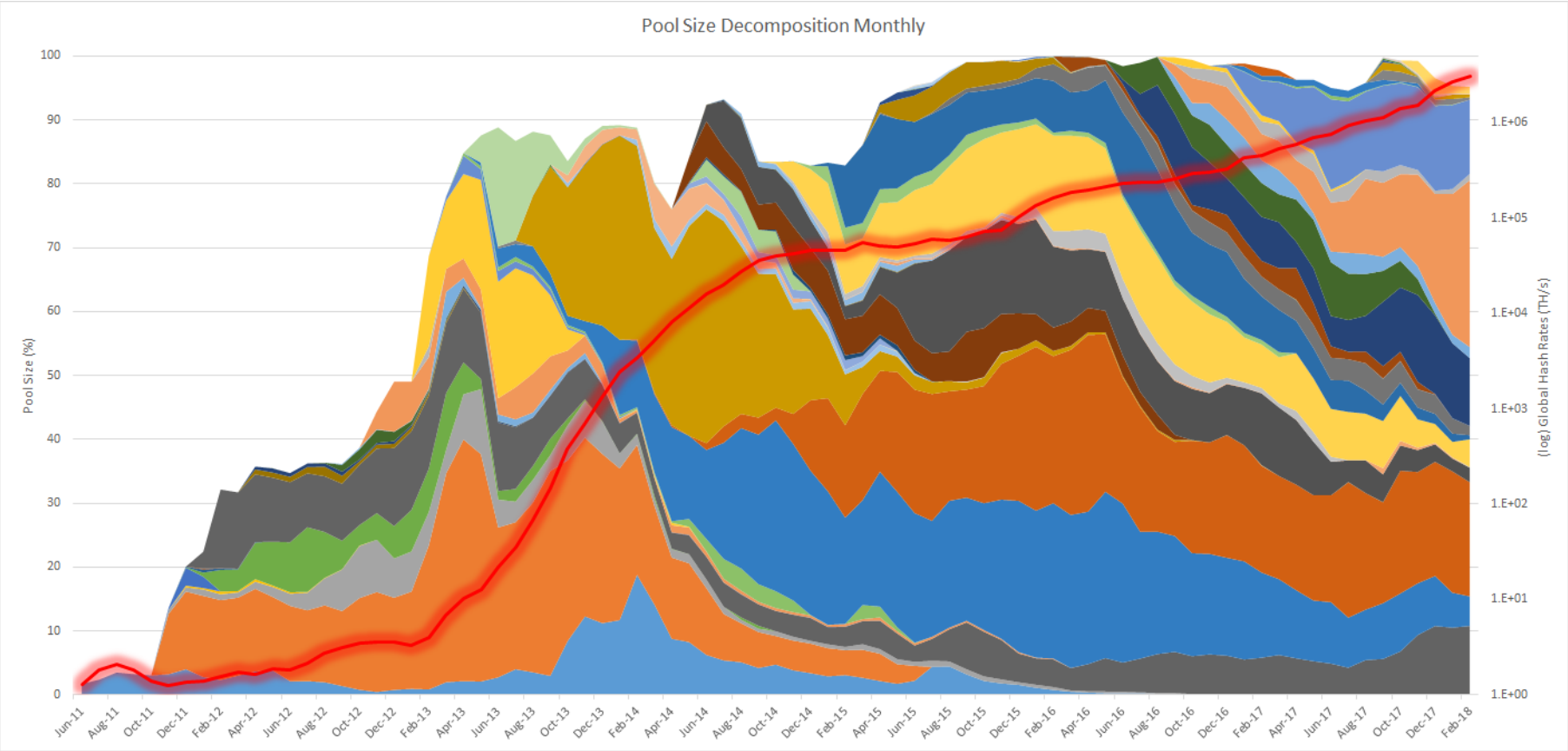
Decentralized Consensus: the Bitcoin Example

- Digital/online transactions & central record-keeper
  - Visa Inc. for credit card transactions, central banks for clearing, etc.
- Bitcoin: a **decentralized** cryptocurrency.
- Generating/maintaining decentralized consensus.
  - Mining and Proof-of-Work (PoW): open tournament for miners (independent computers) with rewards.
    - ★ But, mining is a zero-sum game. Arms race.
  - Rewards only valid if endorsed by subsequent miners $\rightarrow$ honest recording $\rightarrow$ no double-spending.
  - Open access and trustless $\rightarrow$ no single point of failure

Rise of Mining Pools

- Bitcoin's (PoW or other protocols) well-functioning relies on adequate decentralization.
- Decentralization: *technological* possibility vs *economic* reality?
- Miners pool in reality
  - "Pooled mining" completely dominates "solo mining"
  - Concerns over sustainability (51% attack, selfish mining, etc.)
  - We offer some fresh economic analyses

Evolution of Bitcoin Mining



The evolution of Bitcoin mining pool size shares

- hashrates rise with pools...
- pools grow first then slow down...

Preview of Results

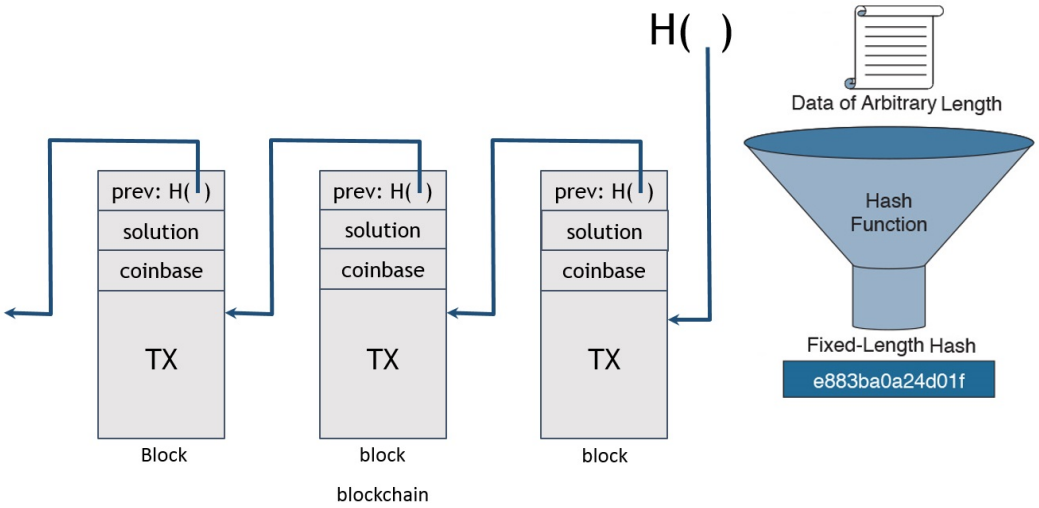
- Risk-aversion $\Longrightarrow$ pooling: significant risk-sharing benefits
  - Diversifying via pools improves (risk-averse) individual payoff but worsens the arms race of mining, quantitatively significant.
  - **Links egregious energy use with pools**; financial innovation improves risk-sharing but aggravates arms race (5$\sim$10 times).
- Risk sharing $\not\Longrightarrow$ pools to merge or grow
  - Under mining technology, miners can join multiple pools to diversify by themselves across pools (M&M insight)!
- An equilibrium model of the mining industry
  - Miners acquire and allocate hash power.
  - Pool owners (enter and) charge fees.
  - Pool's initial passive hash rates as an IO friction, monopolistic competition (robust to entry).
- Empirical evidence from Bitcoin data

Outline

- Introduction
- **Mining Pools**
- Model & Equilibrium
- Empirical Analysis
- Discussion & Conclusion

Bitcoin Mining 101

- Miners repeatedly compete to record recent transactions (aka attaching a block to the chain)
- Winner receives coinbase (currently 12.5BTC) + transactions fees
- A tournament through solving cryptographic puzzles
  - Enumeration (brute force) to find a solution
  - **Hash**(solution, block) has adequate leading zeros
- Difficulty adjustment: 1 block/10 mins on average
  - **The exact source of arms race externality**



Characterizing (Solo) Mining Payoffs

Solution's Poisson arrival: rate proportional to share of global hashrates

- Miner's payoff:

$$X_{solo} = \tilde{B}_{solo} R - c(\lambda_A, T), \text{ where}$$

- $\tilde{B}_{solo} \sim \text{Poisson}\left(\frac{1}{D}\frac{\lambda_A}{\Lambda} T\right)$: # blocks found in T
- Λ: **global hashrate**
- $D = 60 \times 10$ secs: const.
- R: dollar reward per block (coinbase + TX fees).
- $c(\lambda_A, T) = c\lambda_A T$: cost of operation/electricity.

Characterizing (Solo) Mining Payoffs

Solution's Poisson arrival: rate proportional to share of global hashrates

- Miner's payoff:

$$X_{solo} = \tilde{B}_{solo} R - c(\lambda_A, T), \text{ where}$$

- $\tilde{B}_{solo} \sim \text{Poisson}\left(\frac{1}{D}\frac{\lambda_A}{\Lambda} T\right)$: # blocks found in T
- Λ: **global hashrate**
- $D = 60 \times 10$ secs: const.
- R: dollar reward per block (coinbase + TX fees).
- $c(\lambda_A, T) = c\lambda_A T$: cost of operation/electricity.

Rise of Mining Pools

A (proportional) mining pool

- combines multiple miners' hash rates to solve one puzzle
- distributes rewards in proportion to rate contributions

Over T, payoff to a miner with λ_A who joins a (free) pool with λ_B is

$$X_{pool} = \frac{\lambda_A}{\lambda_A + \Lambda_B} \tilde{B}_{pool} R - c(\lambda_A, T), \text{ where}$$

- $\tilde{B}_{pool} \sim \text{Poisson}\left(\frac{1}{D}\frac{\lambda_A+\Lambda_B}{\Lambda} T\right)$

Rise of Mining Pools

A (proportional) mining pool

- combines multiple miners' hash rates to solve one puzzle
- distributes rewards in proportion to rate contributions

Over T, payoff to a miner with λ_A who joins a (free) pool with λ_B is

$$X_{pool} = \frac{\lambda_A}{\lambda_A + \Lambda_B} \tilde{B}_{pool} R - c(\lambda_A, T), \text{ where}$$

- $\tilde{B}_{pool} \sim \text{Poisson}\left(\frac{1}{D}\frac{\lambda_A+\Lambda_B}{\Lambda} T\right)$

Solo vs Pool

A miner with λ_A over period T:

$$X_{solo} = \tilde{B}_{solo} R - c(\lambda_A, T), \tilde{B}_{solo} \sim \text{Poisson}\left(\frac{1}{D}\frac{\lambda_A}{\Lambda} T\right)$$

$$X_{pool} = \frac{\lambda_A}{\lambda_A+\Lambda_B} \tilde{B}_{pool} R - c(\lambda_A, T), \tilde{B}_{pool} \sim \text{Poisson}\left(\frac{1}{D}\frac{\lambda_A+\Lambda_B}{\Lambda} T\right)$$

X_{pool} second-order stochastically dominates X_{solo}, risk-sharing benefit

Illustration of Significant Risk-sharing Benefits

- $\lambda_A = 13.5(\text{TH/s})$: Bitmain Antminer S9 ASIC miner
- $\lambda_B = 3,000,000(\text{TH/s})$: scale of one large mining pool
- $R = \$100,000$ $((12.5+ \sim 0.5)BTC/\text{block} \times \$8\text{K/BTC} \Rightarrow \$104\text{K})$
- CARA $\rho = .00002$ (CRRA of 2 / wealth of \$100K)
- $T = 3600 \times 24\text{s}$: one day.

We have

- $CE_{solo} = \$4.00$ vs $CE_{pool} = \$9.26$, a 131% boost!
- Quantitatively large risk-sharing benefit even for a small pool: $\Lambda_B = 13.5$, about $\sim 20\%$ of boost

Caveat: in our model miners are deciding how to allocate across pools, not whether or not join pools

Evolution of Pool Sizes and Fee Contracts

Year	Hashrate (PH/s) (A)	# of Pools (B)	Top 5 (%) (C)	Avg Fee (Size.W.) (%) (D)	# Frac. Prop (%) (E)	Fee (%)			
						Top 5		All	
						Prop. (F)	Ave. (G)	Prop. (H)	Ave. (I)
2011	0.01	8	7.63	0.57	87.12	0.28	0.28	0.28	0.25
2012	0.02	15	34.66	2.71	61.25	0.66	1.76	0.65	1.56
2013	1.48	23	71.01	2.73	62.57	1.58	2.29	1.16	2.02
2014	140.78	33	70.39	0.88	70.50	1.33	1.13	0.88	2.38
2015	403.61	43	69.67	1.51	77.92	1.10	1.31	0.84	1.33
2016	1,523.83	36	75.09	2.50	77.14	1.48	2.15	0.97	1.67
2017	6,374.34	43	62.25	1.67	78.89	2.00	1.43	1.42	1.32

Outline

- Introduction
- Mining Pools
- **Model & Equilibrium**
- Empirical Analysis
- Discussion & Conclusion

Model Setup

- Static game, CARA $u(x) = \frac{1}{\rho}\left(1 - e^{-\rho x}\right)$
- Measure N active miners acquire hash rate λ_a, taking equilibrium $\{f_m\}_{m=1}^{M}$ as given
  - N large to rule out solo mining.
- Symmetric equilibrium: all active miners same allocation
  - Pools might be heterogeneous with initial sizes
- M pool managers set fees f_m to compete.
- "Friction": pool m endowed with passive hash rates Λ_{pm}
  - e.g. inattentive miners
  - key to monopolistic competition
  - empirical link to initial pool size

Model Setup

- Static game, CARA $u(x) = \frac{1}{\rho}\left(1 - e^{-\rho x}\right)$
- Measure N active miners acquire hash rate λ_a, taking equilibrium $\{f_m\}_{m=1}^{M}$ as given
  - N large to rule out solo mining.
- Symmetric equilibrium: all active miners same allocation
  - Pools might be heterogeneous with initial sizes
- M pool managers set fees f_m to compete.
- "Friction": pool m endowed with passive hash rates Λ_{pm}
  - e.g. inattentive miners
  - key to monopolistic competition
  - empirical link to initial pool size

Active Miner's problem

$$\mathbb{E}\left[u\left(\sum_{m=1}^{M}\left(\frac{\lambda_m \tilde{B}_m(1-f_m)}{\Lambda_{am}+\Lambda_{pm}}\right)R - C\sum_{m=1}^{M}\lambda_m\right)\right] \tag{1}$$

the problem reduces to

$$\max_{\lambda_m \geq 0}\left[\frac{\Lambda_{am}+\Lambda_{pm}}{\rho\Lambda}\left(1-e^{-\frac{\rho R(1-f_m)\lambda_m}{\Lambda_{am}+\Lambda_{pm}}}\right) - C\lambda_m\right], \forall m, \tag{2}$$

where the global hash rate Λ is

$$\Lambda = \sum_{m=1}^{M}(\Lambda_{am}+\Lambda_{pm}). \tag{3}$$

Active Miner's problem

$$\mathbb{E}\left[u\left(\sum_{m=1}^{M}\left(\frac{\lambda_m \tilde{B}_m(1-f_m)}{\Lambda_{am}+\Lambda_{pm}}\right)R - C\sum_{m=1}^{M}\lambda_m\right)\right] \tag{1}$$

the problem reduces to

$$\max_{\lambda_m \geq 0}\left[\frac{\Lambda_{am}+\Lambda_{pm}}{\rho\Lambda}\left(1-e^{-\frac{\rho R(1-f_m)\lambda_m}{\Lambda_{am}+\Lambda_{pm}}}\right) - C\lambda_m\right], \forall m, \tag{2}$$

where the global hash rate Λ is

$$\Lambda = \sum_{m=1}^{M}(\Lambda_{am}+\Lambda_{pm}). \tag{3}$$

Pool Managers' Problem

Given $\{\Lambda_{pm}\}_{m=1}^{M}$ and f_{-m}, manager m with fee f_m has a cashflow of

$$\tilde{B}_{pool,m} \cdot Rf_m, \text{ with } \tilde{B}_{pool,m} \sim \text{Poisson}\left(\frac{1}{D}\frac{\Lambda_{am}+\Lambda_{pm}}{\Lambda}T\right)$$

Any pool owner's problem becomes

$$\max_{f_m} \frac{\Lambda_{am}(f_m)+\Lambda_{pm}}{\rho\Lambda(f_m,f_{-m})}\left(1-e^{-\rho R f_m}\right). \tag{4}$$

- Managers take into account the effect of their own fees $\{f_m\}_{m=1}^{M}$ on global hashrates Λ;
- .....infinitesimal miners do not.

Pool Managers' Problem

Given $\{\Lambda_{pm}\}_{m=1}^{M}$ and f_{-m}, manager m with fee f_m has a cashflow of

$$\tilde{B}_{pool,m} \cdot Rf_m, \text{ with } \tilde{B}_{pool,m} \sim \text{Poisson}\left(\frac{1}{D}\frac{\Lambda_{am}+\Lambda_{pm}}{\Lambda}T\right)$$

Any pool owner's problem becomes

$$\max_{f_m} \frac{\Lambda_{am}(f_m)+\Lambda_{pm}}{\rho\Lambda(f_m,f_{-m})}\left(1-e^{-\rho R f_m}\right). \tag{4}$$

- Managers take into account the effect of their own fees $\{f_m\}_{m=1}^{M}$ on global hashrates Λ;
- .....infinitesimal miners do not.

Equilibrium Definition

Equilibrium Definition

A symmetric equilibrium is a collection of $\{f_m\}_{m=1}^M$ and $\{\lambda_m\}_{m=1}^M$ so that

- **Optimal fees:** $\{f_m\}_{m=1}^M$ solves each manager's problem
- **Optimal hash rates allocation:** given $\{f_m\}_{m=1}^M$, $\{\lambda_m\}_{m=1}^M$ solve each active miner's problem
- **Market clearing:** $\Lambda_{am} = N\lambda_m$

- initial size distribution $\{\Lambda_{pm}\}_{m=1}^M$, resulting size distribution $\{\Lambda_{am} + \Lambda_{pm}\}_{m=1}^M$. Pool growth $\frac{\Lambda_{am}}{\Lambda_{pm}}$

A Frictionless Benchmark: $\Lambda_{pm} = 0$

Proposition (Irrelevance of Pool Size Distribution)

- $f_m = 0$ *for all* m *(a Bertrand insight)*
- *any allocation* $\{\lambda_m\}_{m=1}^M$ *with* $\Lambda = N\sum_{m=1}^M \lambda_m = \frac{R}{C}e^{-\rho R/N}$.

- Miners have perfect risk sharing by themselves.
- M$M: why a larger pool when individuals can diversify freely?
  - Fallacy of "risk-diversification $\Longrightarrow$ pools merge/centralization"

- Dark side of pools: marginal benefit of $\frac{R}{C}e^{-\rho R/N}$ with full risk-sharing, v.s. $\Lambda = \frac{R}{C}e^{-\rho R}$ with solo.

Equilibrium with Passive Hash Rates

Active miner's FOC:

$$\underbrace{\frac{R(1-f_m)}{\Lambda}}_{\text{risk-neutral valuation}} \underbrace{e^{-\rho R(1-f_m)\frac{\lambda_m}{\Lambda_{am}+\Lambda_{pm}}}}_{\text{risk aversion discount}} = \underbrace{C}_{\text{marginal cost}} . \tag{5}$$

- Like monopolistic competition (when $\lambda_m = 0$, marginal benefit = risk-neutral valuation).
- Larger pools attract more allocation.

In equilibrium $N\lambda_m = \Lambda_{am}$. Hence

$$\frac{\lambda_m}{\Lambda_{pm}} = \max\left\{0, \frac{\ln\frac{R(1-f_m)}{C\Lambda}}{\rho R(1-f_m) - N\ln\frac{R(1-f_m)}{C\Lambda}}\right\} \tag{6}$$

Equilibrium with Passive Hash Rates

Active miner's FOC:

$$\underbrace{\frac{R(1-f_m)}{\Lambda}}_{\text{risk-neutral valuation}} \underbrace{e^{-\rho R(1-f_m)\frac{\lambda_m}{\Lambda_{am}+\Lambda_{pm}}}}_{\text{risk aversion discount}} = \underbrace{C}_{\text{marginal cost}} . \tag{5}$$

- Like monopolistic competition (when $\lambda_m = 0$, marginal benefit = risk-neutral valuation).
- Larger pools attract more allocation.

In equilibrium $N\lambda_m = \Lambda_{am}$. Hence

$$\frac{\lambda_m}{\Lambda_{pm}} = \max\left\{0, \frac{\ln \frac{R(1-f_m)}{C\Lambda}}{\rho R(1-f_m) - N \ln \frac{R(1-f_m)}{C\Lambda}}\right\} \tag{6}$$

Main Results Overview

Proposition

Same fee, same growth; higher fee, lower growth.

- *if* $f_m = f_{m'}$, *then* $\frac{\Lambda_{am}}{\Lambda_{pm}} = \frac{\Lambda_{am'}}{\Lambda_{pm'}}$;
- *if* $f_m > f_{m'}$ *then* $\frac{\Lambda_{am}}{\Lambda_{pm}} < \frac{\Lambda_{am'}}{\Lambda_{pm'}}$.

Main Results

1. Symmetric pools with $\Lambda_{pm} = \Lambda_p$ for all m, we characterize the equilibrium and study the social cost of mining pools
   - Oligopolistic pools take arms race into account, charge positive fees $\Longrightarrow$ less global hashrates Λ than full risk-sharing but more Λ than solo
2. What if heterogeneous pools: Larger pools charge higher fees?
   - Yes, because larger pools take into account of arms race effect more

Main Results Overview

Proposition

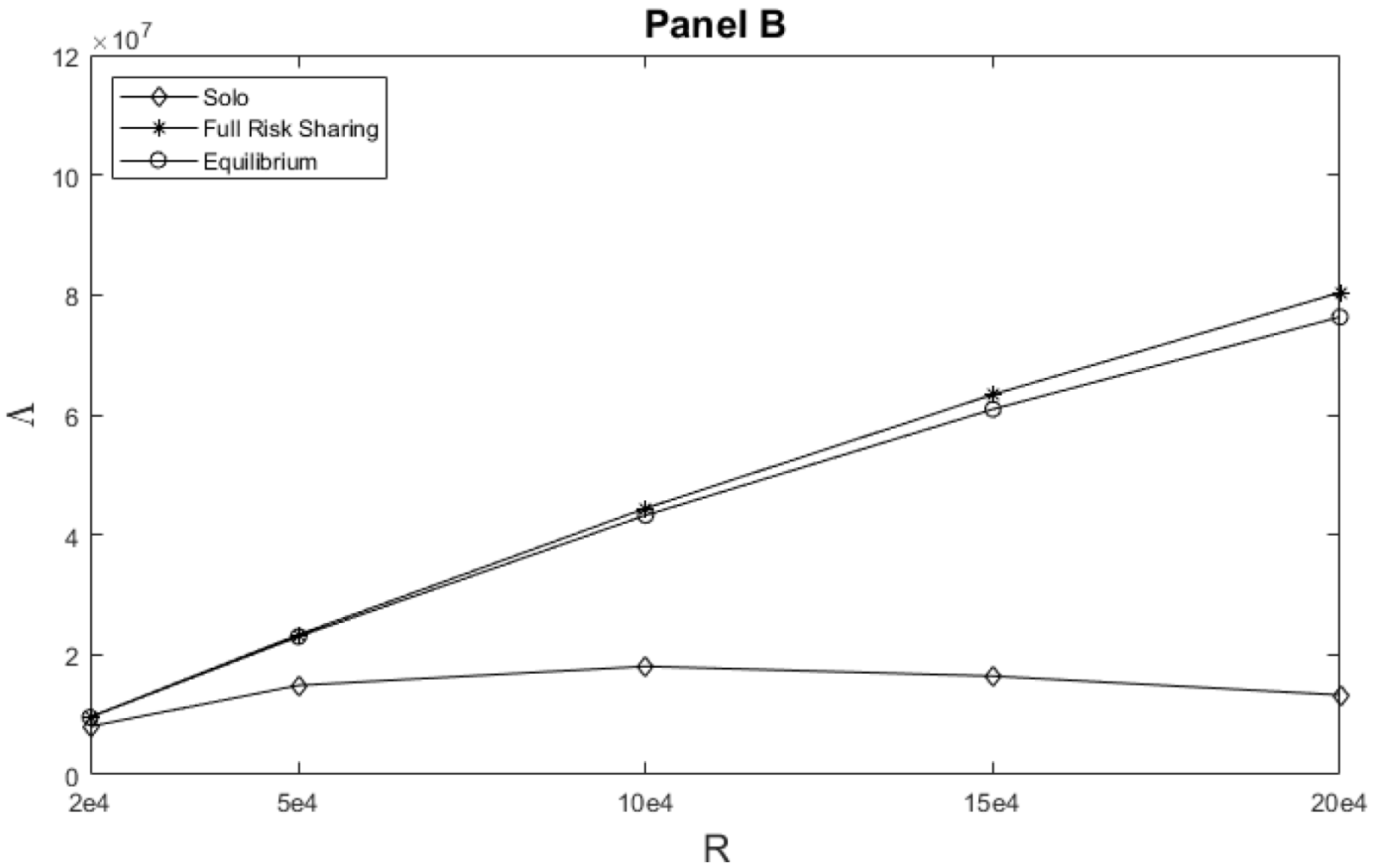
Same fee, same growth; higher fee, lower growth.

- *if* $f_m = f_{m'}$, *then* $\frac{\Lambda_{am}}{\Lambda_{pm}} = \frac{\Lambda_{am'}}{\Lambda_{pm'}}$;
- *if* $f_m > f_{m'}$ *then* $\frac{\Lambda_{am}}{\Lambda_{pm}} < \frac{\Lambda_{am'}}{\Lambda_{pm'}}$.

Main Results

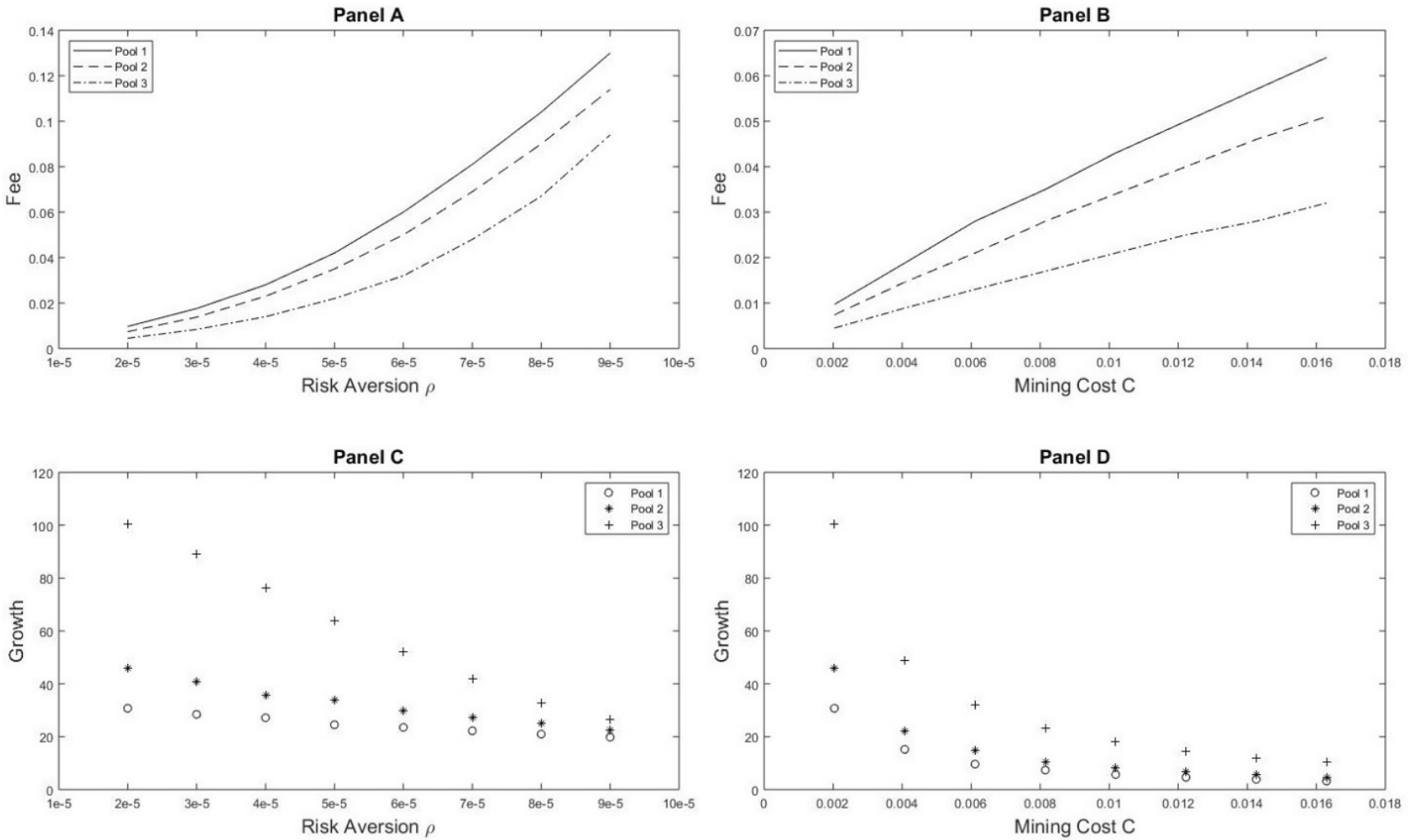
1. Symmetric pools with $\Lambda_{pm} = \Lambda_p$ for all m, we characterize the equilibrium and study the social cost of mining pools
   - Oligopolistic pools take arms race into account, charge positive fees $\Longrightarrow$ less global hashrates Λ than full risk-sharing but more Λ than solo
2. What if heterogeneous pools: Larger pools charge higher fees?
   - Yes, because larger pools take into account of arms race effect more

Social Cost of Mining Pools



$R = 1 \times 10^5$, $N = 10$, $M = 2$, $C = 0.00204$, and $\rho = 1 \times 10^{-5}$.

Pool Evolution: Larger Λ_{pm}, Lower $\frac{\Lambda_{am}}{\Lambda_{pm}}$



$R = 1 \times 10^5$, $\lambda_a = 5 \times 10^4$, $N = 10$, $\Lambda_{p1} = 5 \times 10^5$, $\Lambda_{p2} = 3 \times 10^5$, $\Lambda_{p3} = 1 \times 10^5$, $C = 0.00204$, and $\rho = 2 \times 10^{-5}$.

Outline

- Introduction
- Mining Pools
- Model & Equilibrium
- **Empirical Analysis**
- Discussion & Conclusion

Empirical Evidence: Data and Methodology

Data on pool size (i.e., hashrate share) evolution

- estimated from block relaying records (monthly)
- the newly mined blocks divided by total blocks mined globally

Data on pool fee/reward type evolution

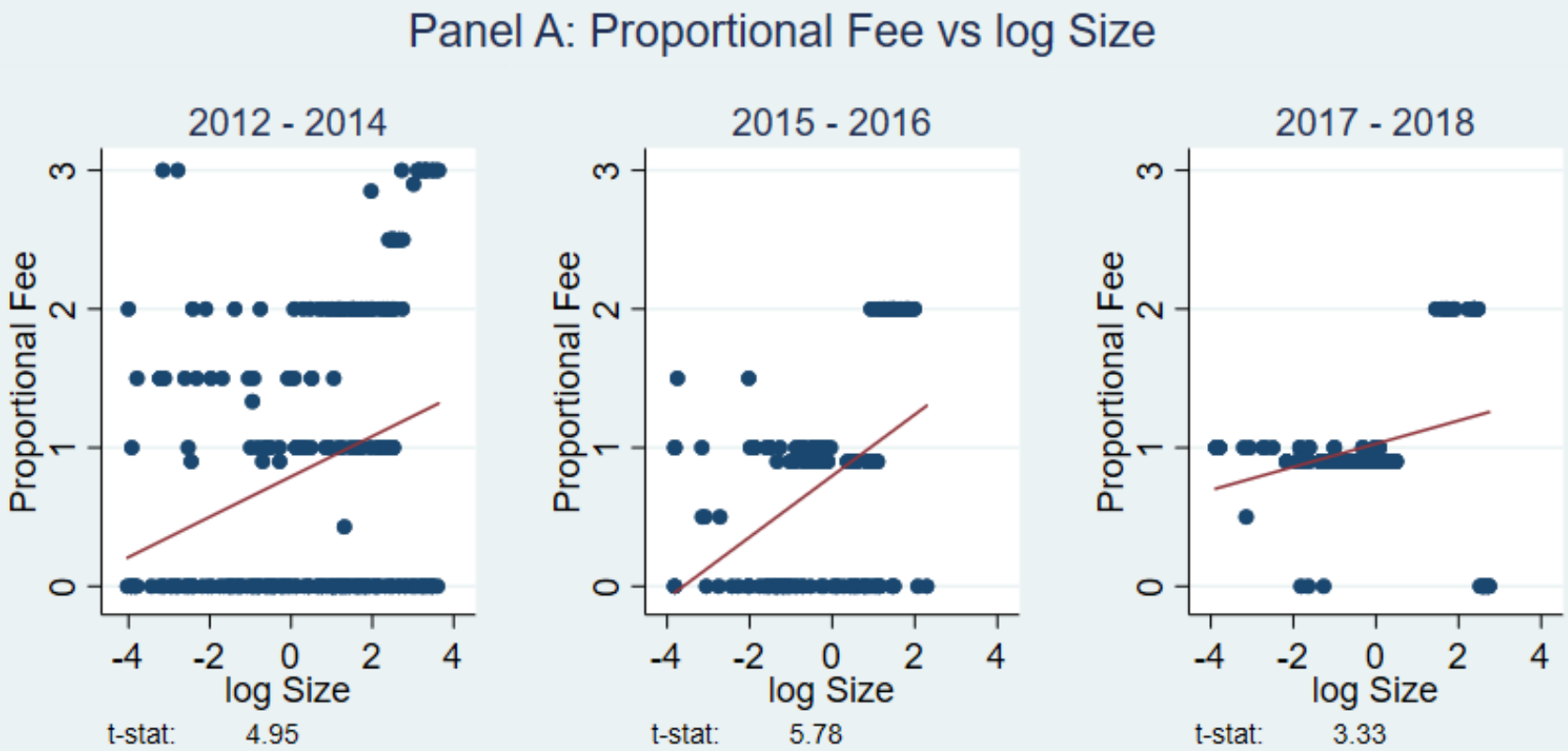
- Bitcoin Wiki: Comparison of mining pools
- the entire Wiki revision history

What we do

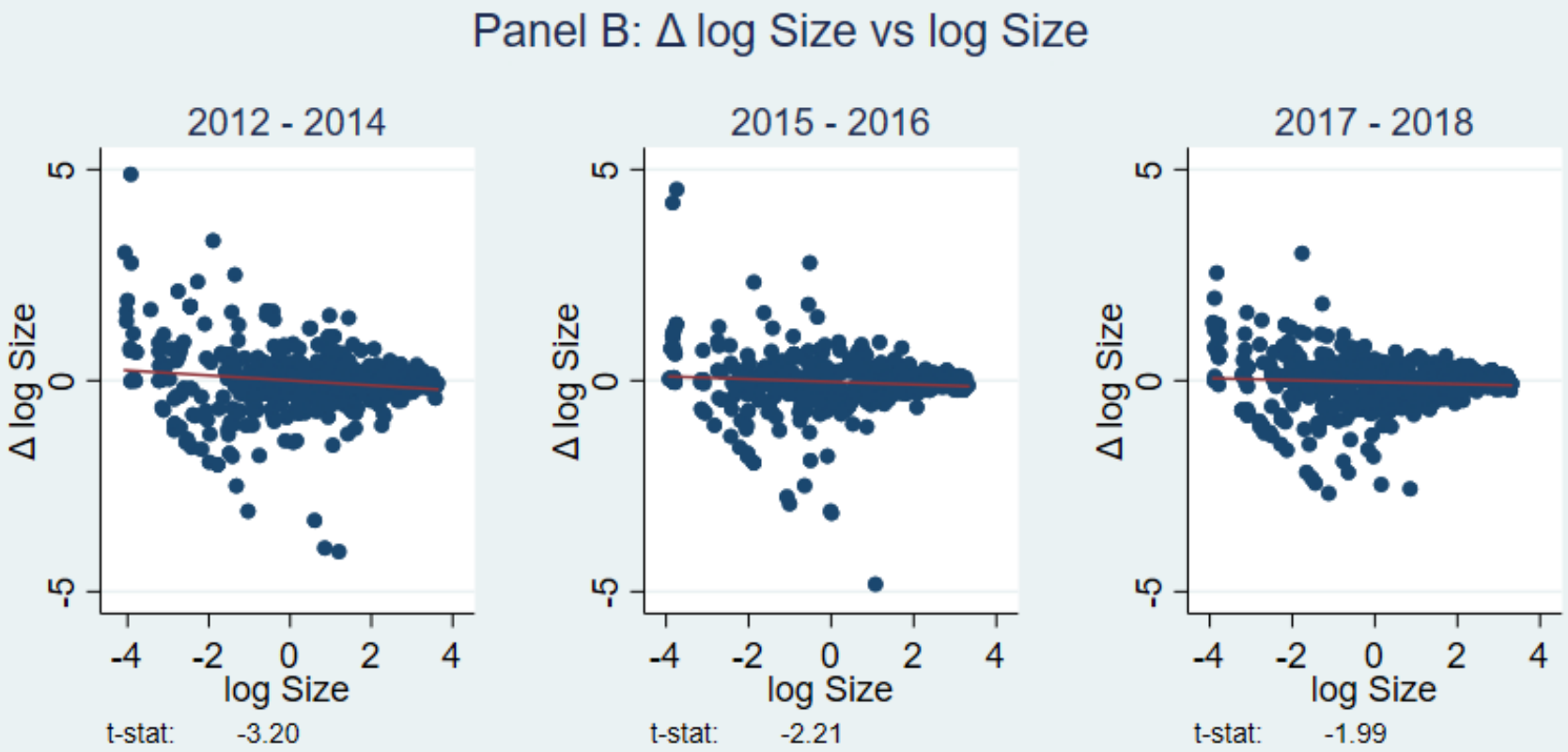
1. investigate relationships between monthly growth rates /average fees and previous month hashrate share in three windows

   (i.e., 2012-2014, 2015-2016, and 2017-2018)

Empirical Evidence: Results



Empirical Evidence: Results



Pool Size, Fee, and Growth: Regression Results

	Panel A: *Proportional Fee*			
	2012-2014	2015-2016	2017-2018	2012-2018
	(1)	(2)	(3)	(4)
logSize	0.16***	0.24***	0.09***	0.16***
	(4.95)	(8.63)	(4.18)	(7.67)
Adjusted R^2	-0.007	0.078	-0.052	-0.002
Month FE	Yes	Yes	Yes	Yes
Observations	286	147	140	573

	Panel B: Δ*logSize*			
	2012-2014	2015-2016	2017-2018	2012-2018
log *Size*	-0.05**	-0.03*	-0.02	-0.03***
	(-2.35)	(-1.90)	(-1.36)	(-3.23)
Adjusted R^2	0.013	-0.004	0.031	0.016
Month FE	Yes	Yes	Yes	Yes
Observations	499	562	644	1705

t statistics in parentheses

$^{*}: p < 0.10, ^{*}* : p < 0.05, ^{*}** : p < 0.01$

Measuring Passive Sizes

1. Identify pool manager addresses from coinbase transactions
   - label all transactions sent from pool manager addresses as paychecks
2. Within each pool, define
   - loyalty addresses: ones having only appeared in a unique pool manager's paychecks
   - seed addresses: top 10 addresses receiving the most bitcoins from the pool manager within a month
   - relationship addresses: top 10% addresses receiving the most bitcoins from the pool manager within a month
3. A pool's loyalty (seed, relationship) size: scale by global hashrates

Loyalty, seed, and relationship sizes are noisy proxies for passive size

Passive Size, Pool Fee, and Growth: Regression Results

	Panel A: *Proportional Fee*			
	log Pool Size	log Loyalty Size	log Seed Size	log Relationship Size
	(1)	(2)	(3)	(4)
Coefficient	0.16***	0.12***	0.17***	0.20***
t statistics	(7.67)	(8.17)	(6.23)	(10.19)
Adjusted R^2	-0.002	-0.077	-0.096	0.013
Monthly FE	Yes	Yes	Yes	Yes
# Obs.	573	396	413	413
	Panel B: $\Delta logSize$ or $\Delta Active\_Growth$			
Coefficient	-0.03***	-9.73***	-0.36***	-0.34***
t statistics	(-3.23)	(-20.49)	(-11.66)	(-16.21)
Adjusted R^2	0.016	0.429	0.128	0.170
Monthly FE	Yes	Yes	Yes	Yes
# Obs.	1705	1154	1287	1287

t statistics in parentheses

*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

Outline

- Introduction
- Mining Pools
- Model & Equilibrium
- Empirical Analysis
- **Discussion & Conclusion**

Conclusion

1. A theory of mining pools
   - Risk-sharing as a natural centralizing force.
   - Financial innovation/vehicle that improve risk-sharing aggravates mining arms race, contributing to egregious energy consumption.
2. Risk-diversification sustains decentralization
   - MM insight, IO insight $\rightarrow$ Blockchain sustainability.
   - Same force, other factors can be added.
   - Empirical evidence: Bitcoin mining industry structure.
3. Theory
   - IO of crypto-mining/consensus generation markets.
   - FinTech/gig/sharing economy; decentralized systems.
   - Monopolistic competition with risk aversion and externality.

Conclusion

1. A theory of mining pools
   - Risk-sharing as a natural centralizing force.
   - Financial innovation/vehicle that improve risk-sharing aggravates mining arms race, contributing to egregious energy consumption.
2. Risk-diversification sustains decentralization
   - MM insight, IO insight $\rightarrow$ Blockchain sustainability.
   - Same force, other factors can be added.
   - Empirical evidence: Bitcoin mining industry structure.
3. Theory
   - IO of crypto-mining/consensus generation markets.
   - FinTech/gig/sharing economy; decentralized systems.
   - Monopolistic competition with risk aversion and externality.