# Decentralized Mining in Centralized Pools *

Lin William Cong[†]; Zhiguo He[‡]; Jiasun Li[§].

July 4, 2018

**Abstract**

An open blockchain's well-functioning relies on adequate decentralization. However decentralization cannot be taken for granted with the rise of mining pools in many presumably distributed cryptocurrency-mining activities, which calls into question the stability and viability of the system. We study the centralization and decentralization forces in the creation and competition of mining pools: risk-sharing benefits attract independent miners to pools, leading to centralization; however, pool concentration can be moderated through cross-pool diversification and endogenous pool fees. In particular, we show that larger pools charge higher fees, leading to disproportionally less miners to join and thus a slower pool size growth. Empirical evidence from Bitcoin mining supports our model predictions.

**JEL Classification:** D47, D82, D83, G14, G23, G28

**Keywords:** Bitcoin, Blockchain, Boundary of the Firm, Cryptocurrency, Decentralization, FinTech, Mining Pools, Risk-Sharing.

*Incomplete and some references missing. We thank Foteini Baldimtsi, Joseph Bonneau, and Tom Ding for helpful discussions; Xiao Yin and Xiao Zhang provided excellent research assistance. We thank seminar participants at Princeton, Baruch, NYU Stern, and George Mason for helpful comments and discussions. The authors gratefully acknowledge funding from the Center of Initiative on Global Markets and the Stigler Center at the University of Chicago Booth School of Business, and from the Multidisciplinary Research (MDR) Initiative in Modeling, Simulation and Data Analytics at George Mason University.

[†]University of Chicago Booth School of Business. Email: will.cong@chicagobooth.edu

[‡]University of Chicago Booth School of Business and NBER. Email: zhiguo.he@chicagobooth.edu

[§]George Mason University School of Business. Email: jli29@gmu.edu

# 1 Introduction

Digital transactions traditionally rely on a central record-keeper, who is trusted to behave honestly and be sophisticated enough to defend against cyber-vulnerabilities. A blockchain instead decentralizes record-keeping, with the best-known application being the P2P payment system Bitcoin (Nakamoto (2008)).[1] Most extant blockchains rely on variants of the proof-of-work (PoW) protocol, often known as "mining," in which independent computers ("miners") dispersed all over the world spend resources and compete repeatedly for the right to record new blocks of transactions.[2] The winner in each round of competition is typically rewarded with some native crypto-tokens, and in the case of Bitcoin, (newly minted) bitcoins. Miners have incentives to honestly record transactions because their rewards are only valid if their records are endorsed by follow-up miners. Compared to a centralized system, blockchain is robust to cyber-attacks as decentralization removes any single point of failure;[3] it is also presumably less vulnerable to misbehaviors and monopoly powers as it shifts the trust on the stewardship of a central book-keeper to the selfish economic incentives of a large number of competitive miners.

But decentralization cannot be taken for granted. Bitcoin only allows decentralization as a *technological* possibility, but it does not guarantee decentralization as an *economic* reality. While Nakamoto (2008) envisions a perfect competition among independent computer nodes across the world, overtime Bitcoin witnesses the rise of mining pools. In "pooled mining", miners partner together and share mining rewards, as opposed to "solo mining" where each miner bears all her own mining risks. From the perspective of economists, this is natural, as partnerships/cooperatives offer the most common organization forms in humans history in achieving risk diversification among individual agents (e.g., the insurance industry). As a result, over time some pools gain significant share of global hash rates (a measure of computation power), with the mining pool GHash.io briefly reached more than 51% of global hash rates in July, 2014.

These observations, together with other relevant centralizing forces, lead to concerns over whether a system under PoW can be stable and can sustain decentralization in the long run. Should we worry about a winner-take-all situation in mining pool development? Would

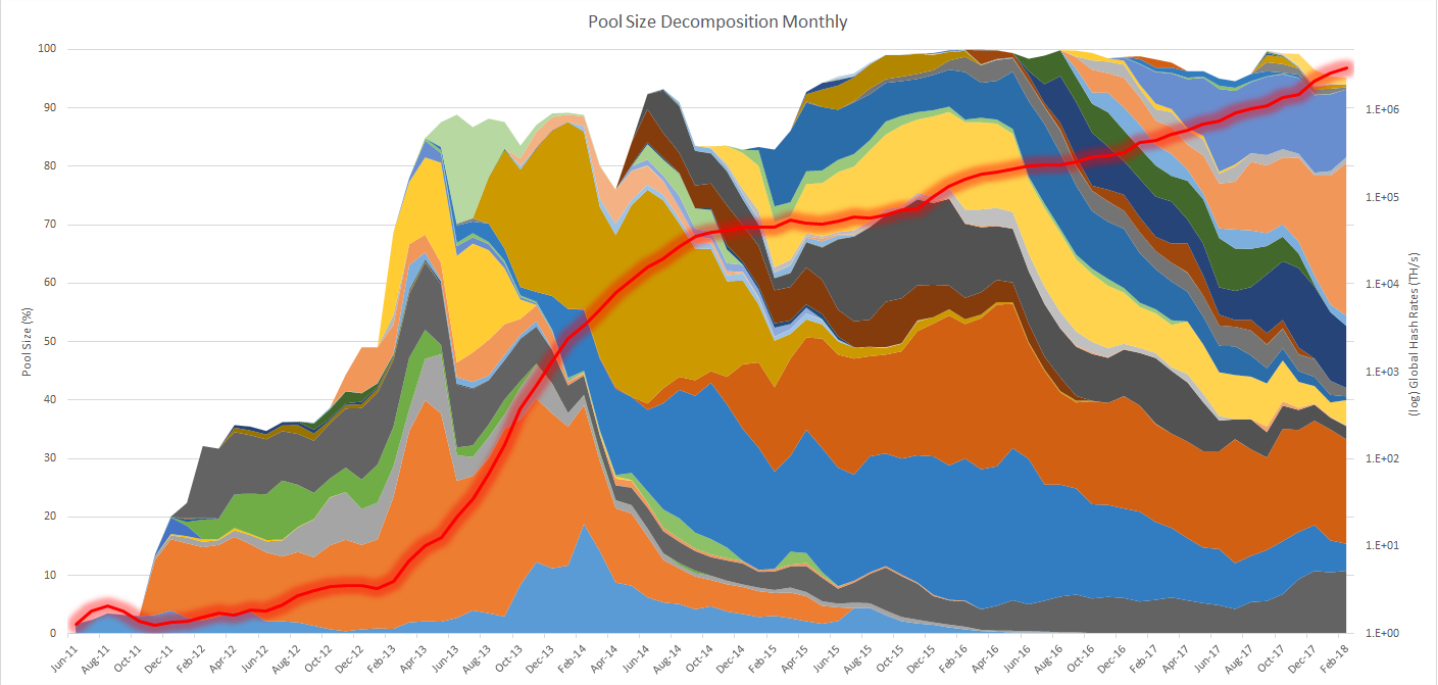
[1] Many retailers already accept Bitcoins (Economist (2017b)). Applications beyond payment systems include the Ethereum platform that enables decentralized smart contract execution. Nanda, White, and Tuzikov. (2017) and Weiss and Corsi (2017) provide a concise introduction to blockchains and applications.

[2] Other protocols for decentralized consensus include Proof-of-Stake (PoS), Practical Byzantine Fault Tolerance (PBFT), etc. Saleh (2017) discusses the sustainability of PoS, among others.

[3] Recent cyber scandals at Equifax offers a vivid lesson. See, e.g., Economist (2017a).

Figure 1: **The evolution of size percentages of Bitcoin mining pools**

This graph plots 1) the growth of aggregate hash (right hand side vertical axis, in log scale) rates starting June 2011 to today; and 2) the size evolutions of all Bitcoin mining pools (left hand side vertical axis) over this period, with pool size measured as each pool's hash rates as a fraction of global hash rates. Different colors indicate different pools, and white spaces indicate solo mining. Over time, Bitcoin mining has been increasingly dominated by mining pools, but no pool seems ever to dominate the mining industry. The pool hash rates information comes from Bitcoinity and BTC.com). For more details, see Section 3.5.




a decentralized record-keeping system be just another Utopian wishful thinking? These questions are not only of interest to the blockchain community, but also fundamental to our understanding of the trade-offs involved in decentralized versus centralized systems (e.g., Hayek (1945)).

Figure 1 illustrates the evolution of the distribution of hash rates among Bitcoin mining pools. Clearly, overtime solo mining has been marginalized and mining pools gradually dominate: mining pools represented less than 5% of the global hash rates at the beginning of June 2011 but has occupied almost 100% since late 2015. This phenomenon suggests natural economic forces toward centralization within a supposedly decentralized system. On the other hand, the equally interesting fact is, while large pool do arise from time to time, none of them have snowballed to dominate global mining. Indeed, there seems to be a mean reverting tendency for large pools. This observation hints at concurrent economic forces that suppress over-centralization.

To better understand if a blockchain system could sustain adequate level of decentralization in the long run, we need a comprehensive understanding of the multiple economic force at play in shaping the industrial organization of bitcoin mining. To this end, our paper attempts shed lights on a non-exhaustive list of economic questions: (1) How large is the risk diversification benefit that mining pools are offering that leads to their rise in the first place? (2) What forces (if any) fuel the continued growth of a mining pool? (3) Are there economic forces that counteract the centralization tendency? and (4) How centralized would pool size distribution be in equilibrium?

Specifically, we model the decision making of miners in allocating their hash rates into mining pools, together with pool managers who charge fees as compensations for offering risk-diversification services. With respect to question (1), we compare a miner's expected utility between solo mining and that from joining a mining pool. Given standard risk aversion calibrations and realistic mining technology parameters, we find that the risk diversification benefit of joining a pool, relative to solo mining, is quantitatively huge: the certainty equivalent of joining a large pool more than doubles that from solo mining.

For question (2), we find that absent other considerations, a larger pool offers higher risk sharing benefits. This is a well-understood economic force: akin to the situation for insurance companies, diversification works better when the firm covers a bigger market.

While this may lead to the hasty conclusion that a large pool will get ever larger, we demonstrate that the risk-sharing benefit *within* a large pool could be alternatively obtained through miner diversification *across* multiple pools. This finding allays concerns that larger pools would grow disproportionally larger, and is reminiscent of the Modigliani-Miller insight that should be well-known to economists: Although investors are risk-averse, firms should not form conglomerate for risk diversification purpose, simply because investors can diversify by themselves in the financial market (by holding a diversified portfolio). Formally, we show that in our model in a frictionless benchmark with risk averse agents, the pool size distribution–just like firm size distribution in the Modigliani-Miller setting–is irrelevant.

Based on this observation, we incorporate into our framework an empirically relevant friction: in this economy, some miners do not optimally diversify, whom we term "passive miners." Otherwise we keep the standard model structure in economics, i.e., multiple pool managers choose fees to compete for customer miners to maximize profits; facing these fees active miners optimally allocate their hash rates across these pools.

We fully characterize the equilibrium in this static setting, and find that the initial pool size distribution matters for welfare and future evolution. In equilibrium, a large proportional

pool optimally charges a high fee which slows its continued growth. In other words, if we put our model into a dynamic setting, pool sizes mean revert endogenously.

The central force of our model is that the larger pool who is offering better risk-diversification service would like to charge a higher fee, hence attracting less active miners to join and slowing down the growth. In terms of comparative statics, when the initial pool size distribution is more disparate, larger pools charge higher fees, shunning away some active miners they could have gotten with a lower fee and hence their sizes mean-revert faster. When the economy becomes more risk averse, larger pools who are offering better diversification service are charging a higher fee and attracting more miners (and hence slower mean-reverting speed).

These theoretical results speak to questions (3) and (4) that, absent other considerations, we should expect an oligopoly market structure of the global mining industry to sustain in the long run, and no single pool will grow too large to monopolize Bitcoin mining. Our theory is closely linked to the risk-sharing benefit that gives rise to mining pools in the first place. Other external forces the blockchain community recognizes that also counteract the natural centralizing force of risk-sharing (e.g., concerns for DDoS attacks or depreciation of coin values) could be added upon our framework, and are discussed in Section 4.

Empirical evidence supports our theoretical predictions. Every quarter, we sort pools into deciles based on the start-of-quarter pool size, and calculate the average pool share, average fee, and average log growth rate for each decile. We show that pools with larger start-of-quarter size charge higher fees, and grow slower in percentage terms. We investigate these relationship in three two-years spans (i.e., 2012-2013, 2014-2015, and 2016-2017), and find almost of them are statistically significant with the signs predicted by our theory.

**Related Literature**

Our paper contributes to emerging studies on blockchains and distributed ledger systems. Harvey (2016) briefly surveys the mechanics and applications of crypto-finance, especially Bitcoin. Yermack (2017) evaluates the potential impacts of the blockchain technology on corporate governance. Raskin and Yermack (2016) push further to envision that the central banks might use the technology to launch their own digital currencies. Cong and He (2018) examine decentralized consensus generation on blockchains, and how its direct tension with information distribution affect competition and industrial organization. Among early studies on how the impact of the technology, Malinova and Park (2016) analyze the design of the mapping between identifiers and end-investors and the degree of transparency of holdings in a blockchain-based marketplace, Khapko and Zoican (2017) argue that blockchains provide

flexibilities that could improve financial settlement, and Cong, Li, and Wang (2018) provide a dynamic pricing framework for cryptocurrencies and highlight the roles of tokens on user adoption.

Our study also directly relates to crypto-currency mining games. Nakamoto (2008) outlines the Bitcoin mining protocol as a well-functioning incentive scheme (under adequate decentralization). Biais, Bisiere, Bouvard, and Casamatta (2018) extend the discussion in Kroll, Davey, and Felten (2013) to model mining decisions on which chain to append to as coordination games and analyze equilibrium multiplicity. Kiayias, Koutsoupias, Kyropoulou, and Tselekounis (2016) consider a similar problem with explicit specification of states as trees. Dimitri (2017) models mining as a Cournot-type competition and argues that the dynamic difficulty-adjustment mechanism reduces monopoly power. These papers treat miners as risk-neutral, while we examine risk-averse agents and complement by directly modeling the rationales for mining pools to arise.

An adequate level of decentralization is crucial for the security of a blockchain. Nakamoto (2008) explicitly requires no single party to control more than half of global computing power for Bitcoin to be well-functioning (thus the concept of 51% attack).[4] Eyal and Sirer (2014) study "selfish mining" in Bitcoin blockchain in which miners launch block-withholding attacks even with less than half of the global hash rates. This practice distorts miner incentives and hampers blockchain security.[5] Large miners may also be vulnerable to block-withholding attacks against one another, known as miner's dilemma (Eyal (2015)). These papers follow the convention in the computer science literature and only consider one strategic pool and assume all other miners to behave naively, rather than characterizing a full equilibrium outcome as we do.[6] They also refer to a mining pool as a single decision maker, rather than delving into the incentives of pool managers and participants as we model here. Furthermore, all miners are risk neutral in these papers, hence they take pools as given, and do not analyze why pools arise in the first place.

Many papers study contract design in mining pools (see e.g. Rosenfeld (2011), Schrijvers,

[4]Empirically, Gencer, Basu, Eyal, van Renesse, and Sirer (2018) investigate the extent of decentralization by measuring the network resources of nodes and the interconnection among them.

[5]Sapirshtein, Sompolinsky, and Zohar (2015) develop an algorithm to find optimal selfish mining strategies. Nayak, Kumar, Miller, and Shi (2016) (stubborn mining) goes beyond the specific deviation in Eyal and Sirer (2014) and consider a richer set of possible deviating strategies. They conclude that there is no *one-size-fits-all* optimal strategy for a strategic miner.

[6]See Beccuti, Jaag, et al. (2017) for an exception, where they find that the minimum requirement to withhold is decreasing with the number of withholding miners, and increasing the heterogeneity among players reduces the likelihood to withhold.

Bonneau, Boneh, and Roughgarden (2016), and Fisch, Pass, and Shelat (2017)), which typically consider one single pool and analyze various miner incentives such as hopping or withholding given detailed dynamics within a mining round. We differ by abstracting away from the micro-details about mining strategies. Instead, we focus on the contracting relationships among miners and pool managers, and the interaction of multiple pools in an industrial organization framework.

Our results on the the rise of mining pools for risk sharing connects with strands of literature on theory of the firm (e.g. classical studies include Wilson (1968) on syndicates and Stiglitz (1974) on sharecropping, as well as recent studies such as Li (2015) on private information coordination).

The rest of the paper proceeds as follows. Section 2 introduces the institutional details of Bitcoin mining, mining pools and their fees, together with stylized facts about mining pools. Section 3 sets up the model based on the key risk-diversification benefits of mining pools illustrated in Section 2, and then characterizes the equilibrium, before providing corroborating empirical evidence using Bitcoin data. Section 4 discusses model implications and extensions, including other decentralization forces limiting pool size. Section 5 concludes.

# 2 Mining Pools: Background and Principle

This section provides background knowledge of the Bitcoin mining process, analyzes the risk-sharing benefit of mining pools, and introduces typical pool fee contracts.

## 2.1 Mining and Risky Reward

Bitcoin mining is a process in which miners around the world compete for the right to record a brief history (known as block) of bitcoin transactions. The winner of the competition is rewarded with a fixed number of bitcoins (currently 12.5 bitcoins, or ₿12.5), plus any transactions fees included in the transactions within the block.[7] In order to win the competition, miners have to find a piece of data (known as solution), so that the hash (a one-way function) of the solution and all other information about the block (e.g. transaction details within the block and the miner's own bitcoin address) has an adequate number of leading zeros. The minimal required number of leading zeros determines mining difficulty. The mining difficulty dynamically adjusts over time, so that *on average* one solution is found

[7]See Easley, O'Hara, and Basu (2017) and Huberman, Leshno, and Moallemi (2017).

every 10 minutes. Under existing cryptography knowledge, the solution can only be found by brute force (enumeration). Once a miner wins the right to record the most recent history of bitcoin transactions, the current round of competition ends and a new one begins.

Bitcoin mining is hence analogous to gold mining. Just like a gold miner who spends manpower and energy to dig the ground in search of gold, a Bitcoin miner spends computing powers (known as hash rates) and related electricity/cooling/network expenses in search of solutions to some difficult cryptography puzzles; just like a gold miner who only gets paid when he successfully finds the gold, a bitcoin miner only gets paid when he finds a solution. More importantly, just like gold mining is risky, so is bitcoin mining – when luck is not in favor, a miner could continuously spend mining expenditures within a prolonged period without finding a solution and hence remain unpaid.

Technology rules that the probability of finding a solution is not affected by the number of trials attempted. This well-known memoryless property implies that the event of finding a solution is captured by a Poisson process with the arrival rate proportional to a miner's share of hash rates globally. Precisely, given a unit hash cost $c$ and unit dollar award $R$ for each block, the payoff to the miner who has a hash rate of $\lambda_A$ operating over a period $T$ is

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

Here, $\tilde{N}_{solo}$ is number of blocks the miner finds within $T$, $\Lambda$ denotes global hash rate (i.e., the sum of hash rates employed by all miners), $D = 60 \times 10$ is a constant so that on average one block is created every 10 minutes. The Poisson distributed random variable $\tilde{N}$ captures the risk that a miner faces in this mining game.

Because mining is highly risky, miners have strong incentives to find ways to reduce risk. While theoretically there are various ways to reduce risk, a common practice is to have miners mutually insure each other by creating a (proportional) mining pool. The next section describes how such a mining pool works.

## 2.2 Mining Pool and Risk Sharing

A mining pool combines the hash rates of multiple miners to solve one single cryptographic puzzle, and distributes the pool's mining rewards back to participating miners in proportion to their hash rate contributions. Ignore fees that represent transfers among pool members for now. Then, following the previous example, the payoff to one participating

miner with hash rate $\lambda_A$ who joins a pool with existing hash rate $\lambda_B$ is

$$X_{pool} = \frac{\lambda_A}{\lambda_A + \lambda_B} \tilde{N}_{pool} R - c\lambda_A T, \text{with } \tilde{N}_{pool} \sim \text{Poisson}\left(\frac{1}{D}\frac{\lambda_A + \lambda_B}{\Lambda} T\right) \tag{2}$$

For illustration, consider the symmetric case with $\lambda_A = \lambda_B$. Then compared to solo mining, a miner who conducts pooled mining is twice likely to receive mining payouts but half the rewards at each payment. This is just the standard risk diversification benefit, and we have the following proposition.

**Proposition 1.** *$X_{pool}$ second-order stochastically dominates $X_{solo}$, so any risk-averse miner will strictly prefer $X_{pool}$ than $X_{solo}$.*

Hence pooled mining provides a more stable cashflow and reduces the risk a miner faces.

## 2.3 Quantifying Risk-Sharing Benefits of Pooled Mining

The risk-sharing benefit of joining a mining pool can be substantial. To illustrate the magnitude, we calculate the difference of certainty equivalents of solo mining and pooled mining for a typical miner in practice. Throughout we will use preference with Constant-Absolute-Risk-Aversion, i.e., exponential utility, in this paper:

$$u(x) \equiv \frac{1}{\rho}\left(1 - e^{-\rho x}\right) \tag{3}$$

The resulting magnitude will be more or less robust to this utility specification, as we will calibrate the risk-aversion parameter $\rho$ based on the widely-accepted magnitude of the Relative Risk-Aversion coefficient.

The certainty equivalent of solo mining, which is denoted by $CE_{solo}$, can be computed as

$$CE_{solo} \equiv u^{-1}(\mathbb{E}[u(\tilde{X}_{solo})]) = \frac{\lambda_A T}{D\Lambda}\frac{1}{\rho}(1 - e^{-\rho R}) \tag{4}$$

Similarly, the certainty equivalent of joining a mining pool, denoted by $CE_{pool}$, is

$$CE_{pool}(\lambda_B) \equiv u^{-1}(\mathbb{E}[u(\tilde{X}_{pool})]) = \frac{(\lambda_A + \lambda_B)T}{D\Lambda}\frac{1}{\rho}(1 - e^{-\rho R\frac{\lambda_A}{\lambda_A + \lambda_B}}) \tag{5}$$

We highlight that the certainty equivalent to joining the pool dependes on the current pool size ($\lambda_B$), as typically a larger pool offers greater risk diversiciation benefit.

We plug in some reasonable numbers to gauge the magnitude of risk-sharing benefit of joining the pool. Suppose $\lambda_A = 13.5$(TH/s), which is what one Bitmain Antminer S9 ASIC miner (a commonly used chip in the bitcoin mining industry) can offer, $\lambda_B = 3,000,000$(TH/s), which is at the scale of one large mining pool, $R = \$100,000$ (₿12.5 reward + ∼₿0.5 transaction fees per block and \$8000 per BTC gives \$104,000), and $\rho = .00002$ (assuming a CRRA risk aversion of 2 and a wealth of \$100,000 per miner gives a corresponding CARA risk aversion of 0.00002). Take $T = 3600 \times 24$ which is one day. Then it is easy to calculate $CE_{solo} = 4.00216$ and $CE_{pool} = 9.25710$, which implies a difference of 5.25494. In fact, the difference of certainty equivalents between solo mining and pooled mining is about 57% of the expected reward $\mathbb{E}(\tilde{X}_{solo})$, which is about 9.25714 (note, it is almost identical to $CE_{pool}$ due to the large pool size and almost perfect diversification). From another perspective, for a small miner, joining a large pool almost boost his risk-adjusted payoff by more than 131%! [8]

Even for small pools, the risk-sharing benefit can be still quantitatively large. Given other parameters used in the above calculation, but imagine that the miner is considering to join a small mining pool with only one existing miner with a S9 ASIC chip so that $\lambda_B = 13.5$. In this case, the implied difference in certainty equivalents is about 20% of the reward.

The fact that pools offer a large risk-diversification benefit immediately implies that in a competitive market environment, mining pools can potentially charge fee to newly joined miners. The equilibrium fees, which should be lower than the monopolist fees calculated above due to competition, will depend on the industrial organization structure of mining pools. We will develop a model to study this topic shortly.

In practice, there are various form of compensation/fee contract that individual miners accept when joining a mining pool, a topic we turn to now.

[8] Even if we set $\rho = .00001$ which implies a miner with CRRA risk aversion of 2 and is twice richer, joining this large pool increases his risk-adjusted payoff by more than 85%. For more risk-averse miners (e.g. $\rho = .00004$), given the current mining cost parameters, joining a pool could turn a (certainty equivalent) loss into profit. Assuming a \$0.12 per kWh electricity cost, and 1375w/h for S9 (see here), the power consumption is $c = 1.375 \times 0.12/(3600 \times 13.5)$ per TH (or $c = \$3.96/(3600 \times 24 \times 13.5)$ per TH with \$3.96 daily power cost). Assume $\Lambda = 21,000$(TH/s), then $\frac{1}{D\rho}\frac{\lambda_A+\lambda_B}{\Lambda}\left(1 - e^{-\rho R\frac{\lambda_A}{\lambda_A+\lambda_B}}\right) - \lambda_A c = \$6.1 \times 10^{-5}$/s or \$5.3/day, while $\frac{1}{D\rho}\frac{\lambda_A}{\Lambda}\left(1 - e^{-\rho R}\right) - \lambda_A c = -\$2.0 \times 10^{-5}$/s or −\$1.7/day.

## 2.4 Fee Contracts in Mining Pools

Different pools in practice offer different fee structures to its participating miners, which could be roughly categorized into three classes: *Proportional*, *Pay per Share* (PPS), and *Cloud Mining*. Table **??** gives the full list of contracts currently used by major pools, with Appendix **??** offering a full description of different reward types.

We will discuss these classes of compensation fee structure based on the contracting variables and the mapping from the contracting variables to payoffs. Technical details will be mentioned only when it is important to understand the unique feature of contracting in mining pools.

### 2.4.1 Pool Managers and Mining Reward

A mining pool is often maintained by a pool manager, who takes a cut into miners' rewards at payout, known as pool fees which differ across pool contracts. In practice, all miners are subject to the same pool fee when contributing to the same pool under the same contract, independent of the level of their hash rates contributed to the pool. In other words, there is no observed price discrimination in terms of the pool fee charged.

Furthermore, different pools also vary in how they distribute transaction fees in a block. These transaction fees are different from "compensation/fees" that our model is analyzing; as discussed in Section 2.1, the transaction fees are what bitcoin users pay for record-keeping their transactions currently in mempool (but not on the chain yet) by the newly mined block. While most pools keep transaction fees and only distribute the coin reward from newly created block, given the rise of transaction fees recently more pools now also share transactions fees. Our reduced form block reward $R$ encompasses both types of reward.

### 2.4.2 Effectively Observable Hash Rates

All classes of fee contracts effectively use a miner's hash rate as contracting variable. Although in theory a miner's hash rate is unobservable to a remote mining pool, computer scientists have designed ways to approximate it with high precision by counting the so-called *partial solutions*. A partial solution to the cryptographic puzzle, like solution itself, is a piece of data such that the hash of all information about the block has at least another adequate number of leading zeros that is smaller than the one required by solution. A solution, which can be viewed as "the successful trial," is hence always a partial solution, which can be viewed as "any trial." Counting the number of trials so amounts to measuring

the hash rates. Different observed contracts may use and weigh different partial solutions that essentially represent different approximation methods, but these approximations are all proven to be quite accurate.

Crucially, the approximation error between the measured hash rate and the true hash rate can be set to be arbitrarily small without little cost. For economists, if one interpret "mining" as "exerting effort," then the important implication is that the principal (pool manager) can measure the actual hash rate (miner's effort) in an arbitrarily accurate way, making moral hazard issue almost irrelevant. We hence are in a situation exactly opposite to Holmström (1982): All team members' effort inputs are perfectly observable and hence contractible, and the only relevant economic force is risk diversification.

### 2.4.3 Fee Contracts

As mentioned, there are more than 10 types of fee contracts used in mining pools in practice, though we can classify them into three broad classes: proportional, pay per share (PPS), and cloud mining. These contracts differ in how they map each miner's hash rates to his final reward.

Our paper will focus on proportional fee contracts.[9] Under this contract, each pool participant only gets paid when the pool finds a solution and hence a block. The manager of the pool first charge a fraction, denoted by $f$, of the reward $R$ to this block; then the remaining reward $(1-f)R$ are distributed in proportion to each miner's number of partial solutions found (and hence proportional to their actual hash rates). More specifically, the payoff of any miner with hashrate $\lambda_A$ joining a pool with an existing hashrate $\lambda_B$ and a proportional fee $f$ is

$$\frac{\lambda_A}{\lambda_A+\lambda_B}(1-f)\tilde{N}R - c\lambda_A T, \text{ with } \tilde{N} \sim \text{Poisson}(\frac{1}{D}\frac{\lambda_A+\lambda_B}{\Lambda}T) \tag{6}$$

For pay per share (PPS) contract, each pool participant gets paid from the manager a fixed amount immediately after finding a partial solution (again, in proportional to the hash rate). Hence the PPS contract is corresponding to "hourly-based wages;" or, all participating miners are renting their hash rates to the pool. Following the previous example, given a PPS

[9] In practice, the most salient proportional contract is the variant Pay-Per-Last-N-Shares (PPLNS), which instead of looking at the number of shares in the round, instead looks at the last N shares, regardless of round boundaries.

fee $f_{PPS}$, the participating miner's payoff will be simply $r \cdot \lambda_A$ with

$$r = \frac{RT}{D\Lambda}(1 - f_{PPS}) \tag{7}$$

being the rental rate while giving up all the random block reward. As shown, in practice the PPS fee is quoted as a fraction of the expected reward per unit of hash rate (which equals $\frac{TR}{D\Lambda}$). Cloud mining, which essentially says miners rent hash rates from the pool, does exactly the opposite: a miner pays a fixed amount upfront to acquire some hash rate from the pool, and then gets paid as if conducting solo mining.

Our theoretically analysis focus on proportional fees only, though the economic force can be easily adapted to the case of hybrid of proportional and PPS fees. There are two reasons for this modeling choice. First, in practice, about 70% of pools are adopting proportional fees, and 28% pools are using proportional fees exclusively.

The second reason is more conceptually important. Notice that the pure form of PPS or cloud mining is about risk allocation between miners and pool manager. Under our framework with homogeneous risk aversion among miners and pool managers, there will be no welfare gains by adopting PPS or cloud mining . In contrast, a proportional fee contract embeds the key risk sharing benefit into the contract.

## 2.5 Stylized Facts about Mining Pools

Table 1 serves as a summary of the institutional background of the mining pool industry. With the growth of total hash rates in bitcoin mining (Column $A$), the number of identified mining pools (Column $B$) as well as the concentration of mining pools (Column $C$, measured by C5 which is the total market size of the top-5 pools sorted by hash rates) increase significantly since 2011 but have stayed stagnant around 2016. As a gauge of overall cost in percentage points in joining mining pools, Column $D$ gives the average pool fee (including proportional, PPS, and others) weighted by hash rates for each year. Column E gives the fraction of hash rates in the mining pools that are using proportional fees; though receding in the recent years, in its peak time 2015, this faction is about 75%.

The rest of four columns focus on the evolution and magnitude of pool fees. Column $F$ and $G$ are for top-5 pools while Column $H$ and $I$ for all pools. Overall, the fee falls in the range of a couple of percentage points; and the proportional fees are in general smaller than

Table 1: **Evolution of Pool Sizes and Fees**

This table summarizes the evolution of mining pool sizes and fees from 2011 to 2017. We report total hash rates in Column A, total number of mining pools in Column B, and the fraction of hashrates contributed by top-5 pools in Column C. In Column D, we report the average fee weighted by hashrates charged by mining pools. In Column E, we report the fraction of mining pools that use proportional fees; the fraction is calculated as the number of pools that use proportional fees divided by the number of pools with non-missing information on fee contracts. Column F and G give the average proportional fee and average total fee, both weighted by hashrates, for top-5 pools; and Column H and Column I are for all pools. The pool hash rates information comes from Bitcoinity and BTC.com). The fee contract information is obtained from Bitcoin Wiki. Over time more hash rates are devoted to Bitcoin mining, and a majority of mining pools offer proportional contracts. The largest five pools on average charge higher fees.

| Year | Hashrate (PH/s) (A) | # of Pools (B) | Top 5 (%) (C) | Avg. Fee (Size-Weighted) (%) (D) | # Frac. of Prop. Pools (%) (E) | Fee (%) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Top 5 | | All | |
| | | | | | | Prop. (F) | Ave. (G) | Prop. (H) | Ave. (I) |
| 2011 | 0.01 | 8 | 6.77 | 3.87 | 46.28 | 0.33 | 0.34 | 0.30 | 0.31 |
| 2012 | 0.02 | 15 | 33.2 | 3.48 | 68.57 | 1.81 | 2.09 | 1.26 | 1.60 |
| 2013 | 1.48 | 23 | 60.22 | 3.03 | 67.30 | 1.76 | 2.47 | 1.07 | 1.61 |
| 2014 | 140.78 | 33 | 66.03 | 1.61 | 73.17 | 0.82 | 1.93 | 1.14 | 1.97 |
| 2015 | 403.61 | 43 | 66.29 | 2.24 | 75.14 | 1.50 | 3.13 | 0.78 | 1.32 |
| 2016 | 1,523.83 | 36 | 74.05 | 1.12 | 71.79 | 1.33 | 2.04 | 1.10 | 1.43 |
| 2017 | 6,374.34 | 43 | 54.4 | 0.22 | 63.26 | 1.00 | 1.64 | 0.87 | 1.03 |

"average fee" which is the average of proportional fees, PPS fees, and others.[10]

Last but not least, the stylized fact revealed by comparing "Top 5" and "All" are that fees charged by top 5 pools are higher than the average fees charged by pools with all sizes. This is the empirical pattern that motivates our paper.

# 3 An Equilibrium Model of Mining Pools

We present an equilibrium model where multiple pool managers compete in fees to attract customer miners. We first give a benchmark result: in a frictionless environment where all miners can actively allocate their hash rates, risk-diversification itself does not lead to centralization simply because miners can diversifies by themselves across pools. Pool size distribution starts to matter in an interesting way when we assume that larger pools also have more passive miners who do not adjust their allocations. We show that larger pools

[10]Take PPS fees as an example. As explained, PPS contracts offer zero risk exposure to participating miners, and therefore the miners using PPS contracts are happy to pay a higher fee than using proportional contracts (or equivalently, pool managers will charge more from miners for bearing more risk).

will charge higher fees, leading to slower pool growth. These key theoretical predictions are confirmed in the data.

## 3.1 Setting

Because it is fairly easy for miners to move hash rate contribution across pools, and for mining pools to adjust the fees they charge, we focus on a static model. All agents have the same CARA utility function given in Eq. (3).

There are $M$ mining pools managed by different managers. Pool $m \in \{1, \cdots, M\}$ has $\Lambda_m$ existing hash rates from passive miners who stick to these pools. Empirically, we link $\Lambda_m$ to the pool size, under the assumption that a fixed fraction of miners do not adjust the hash rate contribution to the pool. It could be that they derive alternative benefits from a particular pool, or they do not pay attention to changes in pool sizes or fees. This modeling assumption that only a fraction of players can actively readjust their decisions, in the same spirit of Calvo (1983) is widely used in the literature of dynamic games (e.g., Burdzy, Frankel, and Pauzner (2001) and He and Xiong (2012)).

As explained in Section 2, the mining pool offers significant risk diversification benefit to individual miners. As a result, the manager of pool $m$ charges a (proportional) fee of $f_m$ to maximize her profits, where the fee vector $\{f_m\}_{m=1}^M$ will be determined in equilibrium.

There are $N$ active homogeneous miners, each with an endowed hash rates $\lambda_A$. Taking the fee vector $\{f_m\}_{m=1}^M$ as given, these active miners can allocate their hash rates to the above $m$ pools. Because the number of miners joining each pool is large in practice, we assume $N \geq \rho R$, which is realistic and simplifies our analysis.

**Active Miner's problem** Consider one of the $N$ active miners who faces $\{\Lambda_m\}_{m=1}^M$ and $\{f_m\}_{m=1}^M$. When allocating a hash rate of $\lambda_m$ to pool $m$, the payout from the pool $m$ will be

$$X_{pool}^m = \frac{\lambda_m}{\lambda_m + \lambda_{-m} + \Lambda_m} \tilde{N}_{pool,m}(1 - f_m)R \tag{8}$$

where $\lambda_{-m}$ is the hash rate contribution from other $N-1$ active miners. As a result, the active miner with exponential utility function $u(x) = \frac{1-\exp(-\rho x)}{\rho}$ chooses $\{\lambda_m\}_{m=1}^M$ to maximize

$$\mathbb{E}\left[u\left(\sum_{m=1}^M X_{pool}^m + X_{solo} - c\lambda_A T\right)\right] = \mathbb{E}\left[u\left(\sum_{m=1}^M \left(\frac{\lambda_m \tilde{N}_{pool,m}(1-f_m)}{\lambda_m + \lambda_{-m} + \Lambda_m} + \tilde{N}_{solo}\right) R - c\lambda_A T\right)\right]$$

which is equivalent to solving

$$\max_{\lambda_m} \frac{1}{\Lambda} \left[ \sum_{m=1}^{M} (\lambda_m + \lambda_{-m} + \Lambda_m) \left( 1 - e^{-\frac{\rho R(1-f_m)\lambda_m}{\lambda_m + \lambda_{-m} + \Lambda_m}} \right) + (\lambda_A - \sum_{m=1}^{M} \lambda_m) \left( 1 - e^{-\rho R} \right) \right], \quad (9)$$

where $\Lambda$ is the worldwide total hash rate engaging in mining, $\lambda_m \geq 0$ and $\sum_{m=1}^{M} \lambda_m \leq \lambda_A$. Here, the first term in the square bracket captures the benefit from joining pools, and the second term comes from potential solo mining.

**Pool Managers' problem** The pool manager with the same CARA utility function is setting the proportional fee $f_m$ to maximize her expected utility. A pool manager's (random) payoff from operating the pool with passive hash rate $\Lambda_m$ is $\tilde{N}_{pool,m} R f_m$, , where

$$\tilde{N}_{pool,m} \sim \text{Poisson} \left( \frac{T}{D} \frac{N\lambda_m + \Lambda_m}{\Lambda} \right),$$

and $\lambda_m$ is the hash rate that the pool is able to attract from active miners. Obviously, in equilibrium $\lambda_m$ depends on the fees charged by other mining pools, which is denoted by $f_{-m}$.

We look for the Nash equilibrium in a fee setting game among pools. Given $\{\Lambda_m\}_{m=1}^{M}$ and the fee charged by other pools $f_{-m}$, the $m$-pool manager chooses $f_m$ to maximize

$$\max_{f_m} (N\lambda_m + \Lambda_m) \left( 1 - e^{-\rho R f_m} \right) \quad (10)$$

Notice that $\lambda_m$ is a function of $\{\Lambda_m\}_{m=1}^{M}$ and $\{f_m\}_{m=1}^{M}$ given by the solution to the active miner's problem in (9). The embedded assumption in the above maximization is that the pool is facing an aggregate demand function as a function of the price vector, and all homogeneous active miners are taking symmetric best responses to any potential off-equilibrium price quotes (hence form a symmetric equilibrium in any subgame in off-equilibrium paths).

**Equilibrium definition** We focus on symmetric subgame perfect equilibrium where all homogeneous active miners are taking the same strategies. The equilibrium is a collection of $\{f_m\}_{m=1}^{M}$ and $\{\lambda_m\}_{m=1}^{M}$ so that

(1) Given $\{f_m\}_{m=1}^{M}$, $\{\lambda_m\}_{m=1}^{M}$ solves every active miner's problem in (9);

(2) $\{f_m\}_{m=1}^{M}$ solves pool manager $m$'s problem in (10) for all $m$.

## 3.2 Irrelevance of Pool Distribution in a Frictionless Case

The initial size distribution of mining pools matters because we assume it is proportional to the measure of passive miners $\{\Lambda_m\}_{m=1}^M$. To highlight the role of passive miners in our model, we first analyze the model outcome absent passive miners as a benchmark. In particular, we prove a stark irrelevance result of pool size distribution in the frictionless case where $\Lambda_m = 0$ for all $m$'s.

Even absent passive miners (or *ex ante* pool heterogeneity), active miners can still lead to heterogeneous pool sizes. The first best solution features that all active miners are perfectly diversified among all pools, each having $1/N$ share of each pool. Importantly, the following proposition shows that in the first-best solution and the correspondingly implemented market equilibrium, the individual pool size distribution does not matter.

**Proposition 1** (Irrelevance of Pool Size Distribution)**.** *With $\Lambda_m = 0$ for all $m$'s, any feasible allocation $\{\lambda_m\}_{m=1}^M$ with $\sum_{m=1}^M \lambda_m = \Lambda_A$ for active miners together with zero fees $f_m = 0$ for all $m$ constitute an equilibrium. This equilibrium features every active miner's owning $1/N$ of each mining pools and achieves the first-best allocation. The exact pool size distribution $\{N\lambda_m\}_{m=1}^M$ is irrelevant.*

*Proof.* Under this equilibrium, each active miner's expected utility can be calculated as (taking $\Lambda_m = 0$ in Eq. (9) )

$$\frac{1}{\Lambda}\left[\sum_{m=1}^{M}(\lambda_m + (N-1)\lambda_m)\left(1 - e^{-\rho R\frac{\lambda_m}{\lambda_m+(N-1)\lambda_m}}\right)\right] = \frac{1}{\Lambda}N\lambda_A(1-e^{-\rho R/N}), \tag{11}$$

where we have used the fact that $\sum_{m=1}^M \lambda_m = \lambda_A$, which is the sum of all computational power of active miners in consideration. To show that in this equilibrium the pool managers are charging zero fees, imagine that the manager of pool $m'$ raises its fee to a strictly positive level. In this off equilibrium path with strictly positive fee $f_{m'} = \epsilon > 0$, consider the subgame equilibrium in which all active miners switch their hash rates to other pools without affecting their expected utility in (11) (we still have $\sum_{m=1}^M N\lambda_m = N\Lambda_A$ except that $\lambda_{m'} = 0$). Consequently, no pools are able to make strictly positive profit by charging a positive fee in this equilibrium. Q.E.D. □

Proposition 1 says that as long as mining pools attract all computation powers in consideration (so there is no solo mining), every miner can always achieve the perfect diversification by diversifying his endowed hash rate among all pools. There is no reason for pools to merge

by themselves: joining $m$ pools with proper weights, so that each homogeneous miner owns $1/N$ of each pool, is as if joining a single large pool with the aggregate size of these $m$ pools. From this angle, Proposition 1 reflects the conventional wisdom of a capital market, that in a frictionless market investors can perfectly diversify by themselves, rendering no reason for conglomerates to exist solely for risk sharing.

This insight is thought-provoking given numerous discussions on the centralization implications of risk diversification. In the Bitcoin mining community, media discourse and industry debates have centered on how joining larger pools are attractive and would lead to even more hash rates joining the largest pools, making the pools more concentrated; we revisit this topic in Section 4 when we discuss centralizing forces in decentralized systems. But Proposition 1 clarifies that as long as miners can join the pools in a frictionless way, there is no reason to expect that a single large pool necessarily emerges due to the significant risk diversification benefit it offers. This angle also highlights one key difference between Bitcoin mining pools and traditional firms who do provide valuable insurance to workers against their human capital risks (e.g., Harris and Holmstrom (1982); Berk, Stanton, and Zechner (2010)): in Bitcoin mining industry, it is easy for miners to allocate their computational power across multiple pools, but it is much harder for workers to hold multiple jobs.

## 3.3 A Two-Pool Equilibrium

Before we analyze the $M$-pool problem, it is useful to first consider $M = 2$ to gain the basic intuition of the underlying economic forces. Since we have shown a substantial risk-diversification benefit of joining pools relative to solo mining, we focus on the parameter space that active miners never conduct solo mining in equilibrium (which can be easily checked numerically *ex post*).

### 3.3.1 Equilibrium analysis

In terms of the relationship between a pool's initial size (passive hash rates) and the fee it charges in equilibrium, and between the pool fee and the active miner's allocation of hash rate, we have the following result.

**Proposition 2.** *In an equilibrium whereby active miners only allocate hash rates between two pools (pool 1 and 2),* $\Lambda_1 \geq (>)\Lambda_2$ *implies* $f_1 \geq (>) f_2$ *in equilibrium. Moreover,* $f_1 > f_2$ *implies* $\frac{\lambda_1}{\Lambda_1} \leq \frac{\lambda_2}{\Lambda_2}$.

*Proof.* By contradiction. We only discuss the $\geq$ case as the $>$ case is almost identical.

Suppose otherwise that $\Lambda_1 \geq \Lambda_2$ but $f_1 < f_2$, then no deviations from equilibria give

$$(\Lambda_{A1} + \Lambda_1)\left(1 - e^{-\rho R f_1}\right) \geq \left(\frac{\Lambda_1 \lambda_A}{\Lambda_1 + \Lambda_2} + \Lambda_1\right)\left(1 - e^{-\rho R f_2}\right)$$

$$(\Lambda_{A2} + \Lambda_2)\left(1 - e^{-\rho R f_2}\right) \geq \left(\frac{\Lambda_2 \lambda_A}{\Lambda_1 + \Lambda_2} + \Lambda_2\right)\left(1 - e^{-\rho R f_1}\right),$$

where $\Lambda_{A1}$ and $\Lambda_{A2}$ are the *total* allocation from all active miners to pool 1 and 2 when they charge equilibrium fees $f_1$ and $f_2$, respectively. Notice that $\Lambda_{A1} + \Lambda_{A2} = \lambda_A$, we thus get

$$(\lambda_A + \Lambda_1 + \Lambda_2) \geq (\frac{\Lambda_1 \lambda_A}{\Lambda_1 + \Lambda_2} + \Lambda_1)\frac{1 - e^{-\rho R f_2}}{1 - e^{-\rho R f_1}} + (\frac{\Lambda_2 \lambda_A}{\Lambda_1 + \Lambda_2} + \Lambda_2)\frac{1 - e^{-\rho R f_1}}{1 - e^{-\rho R f_2}}$$

Factoring out $\lambda_A + \Lambda_1 + \Lambda_2$ and multiply $\Lambda_1 + \Lambda_2$ on both sides we have

$$\Lambda_1 + \Lambda_2 \geq \Lambda_1 \frac{1 - e^{-\rho R f_2}}{1 - e^{-\rho R f_1}} + \Lambda_2 \frac{1 - e^{-\rho R f_1}}{1 - e^{-\rho R f_2}},$$

which cannot possibly hold because $f_2 > f_1$ and $\Lambda_1 \geq \Lambda_2$.

The second part is a special case of Proposition 3 and we refer readers to its proof. □

Proposition 2 implies that a (weakly) larger pool charges a (weakly) higher fee. There are two forces at play: First, due to the presence of passive miners, the pool managers consider the benefits of charging a higher fee and getting a higher revenue from the passive miners. This benefit is trivially larger when $\Lambda_m$ is greater. The second force is our emphasis: A larger $\Lambda_m$ attracts active miners because it provides larger diversification benefit. As a result, an active miner may still want to allocate some hash rates to the larger pool who is charging a higher fee.

The result that $\Lambda_1 > \Lambda_2$ leads to $\frac{\lambda_1}{\Lambda_1} \leq \frac{\lambda_2}{\Lambda_2}$ implies that a larger pool will grow slower. Therefore, we see a natural force coming from the industrial organization of mining pools that prevents larger pools from becoming more dominant.

### 3.3.2 Comparative statics and intuition

We investigate the properties of the two-pool equilibrium by studying the comparative statics of the equilibrium fees charged by pool managers $\{f_1, f_2\}$. We are also interested in the equilibrium pool growth $\{\lambda_1/\Lambda_1, \lambda_2/\Lambda_2\}$; recall that under zero solo mining the allocation of hash power to pool 2 is $\lambda_2 = \Lambda_A - \lambda_1$.

Figure 2: **Comparative Statics of Pool Fees and Growth**

Equilibrium fees $\{f_1, f_2\}$ and the growth rate of two pools $\lambda_1/\Lambda_1$ and $\lambda_2/\Lambda_2$ are plotted against miner risk aversion $\rho$ and initial pool size distribution $\Lambda_1/\Lambda_2$, respectively. The baseline parameters are: $R = 1\times10^5$, $\lambda_A = 5\times10^4$, and $N = 50$. In Panel A and C: $\Lambda_1 = 3\times10^6, \Lambda_2 = 1\times10^6$, and $\rho \in [1\times10^{-5}, 3\times10^{-5}]$. In Panel B and D $\rho = 2\times10^{-5}$ and we vary initial pool size distribution $\Lambda_1/\Lambda_2$ while keeping $\Lambda_1+\Lambda_2 = 4\times10^6$.

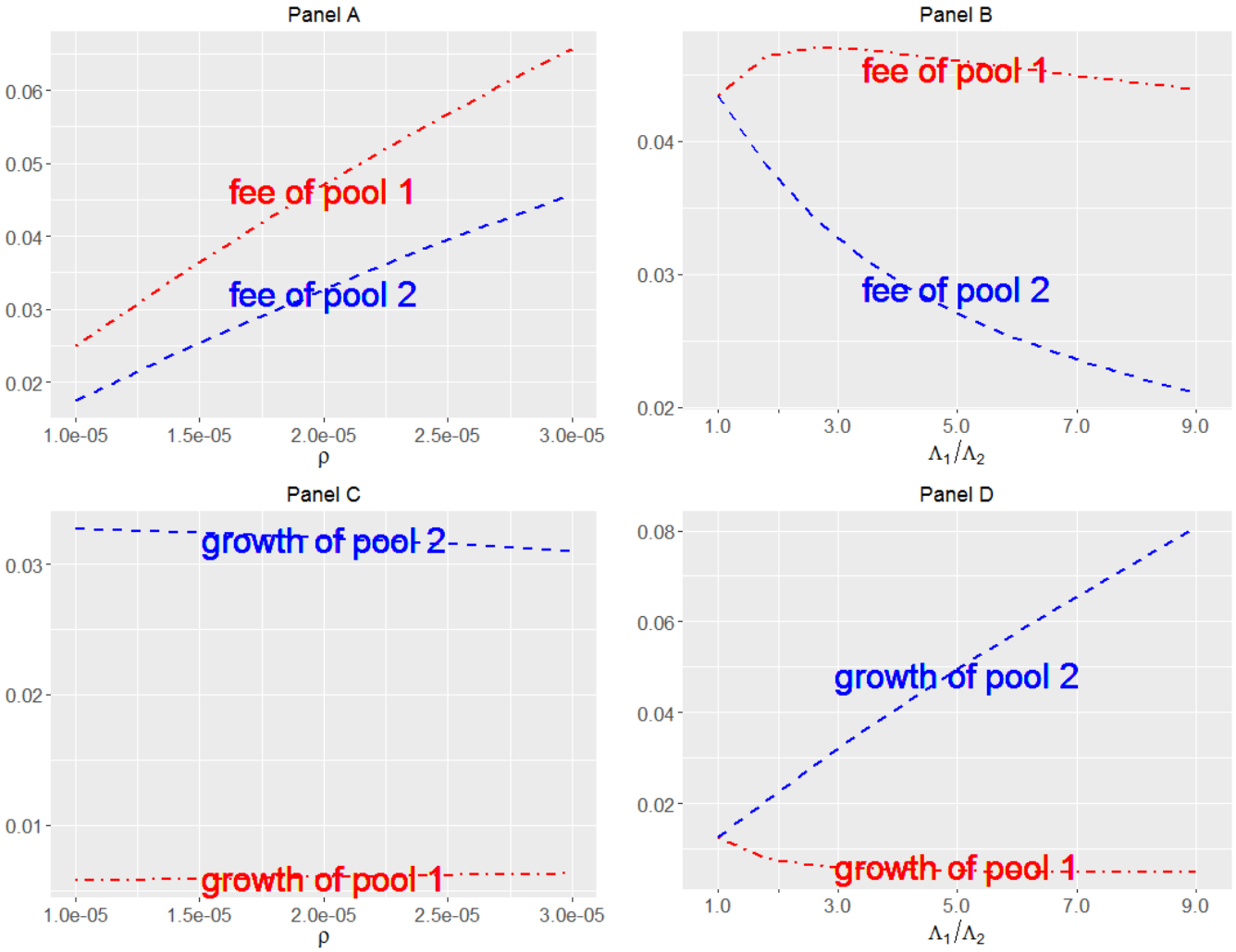



Figure 2 presents the property of our two-pool equilibrium. Without loss of generality, we assume pool 1 has a larger size $\Lambda_1$. Panel A presents how the equilibrium fees respond to exogenous changes in risk aversion $\rho$ in this economy, and Panel B presents how the equilibrium pool growth rates respond to exogenous changes of the distribution of initial pool size $\Lambda_1/\Lambda_2$.

Not surprisingly, when the economic agents becomes more risk averse, mining pools charge higher fees, as show in Panel A. Panel B illustrates how fees change when we vary the

distribution of initial pool size distribution $\Lambda_1/\Lambda_2$ (while keeping $\Lambda_1 + \Lambda_2$ fixed; this way, we are only varying the pool size distribution). As formally shown in Proposition 2, the larger pool is charging a higher fee; and this price differential disappears for two pools with equal initial size when $\Lambda_1/\Lambda_2 = 1$. Starting from there, when pool 1 becomes larger, $f_1$ increases initially while $f_2$ decreases. Eventually when the much smaller pool 2 charges a much lower fee, pool 1 has to lower its fee as well to remain competitive.

Panel C and Panel D concern the growth of two pools. In Panel C, more risk-averse active miners shift their hash power toward the larger pool 1, making the growth of pool 1 (pool 2) to be increasing in the risk-aversion $\rho$. This is because the large pool (with greater passive computing powers) offers greater risk-diversification benefit.

Panel D gives the decentralization force highlighted by this paper. The larger pool 1, because of charging a higher fee, attracts in percentage terms less active computing power compared to the smaller one. Quantitatively, this decentralization force becomes stronger when the wedge of growth rate between the large and small pools diverges. In our model, the decentralization force strengthens when the pool size distribution becomes more unequaled (a higher $\Lambda_1/\Lambda_2$ leads to a greater fee wedge across pools).

## 3.4 Equilibrium Fees and Allocation

Now we analyze the general $M$-pool case. The problem resembles an equilibrium investment problem with externality, and without the usual CARA-Normal assumptions, deriving analytical solutions of fees and allocations is generally infeasible.

Nevertheless, we can analytically characterize how the miner's allocation is related to pool fees in any equilibrium in Proposition 3. Due to the same economic forces that we explained in the previous section with two pools, our numerical solution reveals that the equilibrium pool fee is always increasing in pool size, though we are still working on some formal proof for this statement.

### Equilibrium Relationship between Miner's Allocation and Pool's Fee

Though lacking closed form solution for the active miners' hash rate allocation problem, we can still characterize how pool fees affect the allocation decision of the miners in any symmetric equilibrium with multiple mining pools.

**Proposition 3.** *In any equilibrium with $M$ pools, for any two pools $m$ and $m'$,*

1. *If* $f_m = f_{m'}$ *, then* $\frac{\lambda_m}{\Lambda_m} = \frac{\lambda_{m'}}{\Lambda_{m'}}$*;*

2. *With a large number of miners (i.e.,* $N \geq \rho R$*), if* $f_m > f_{m'}$ *then we have* $\frac{\lambda_m}{\Lambda_m} \leq \frac{\lambda_{m'}}{\Lambda_{m'}}$*. If in addition* $\lambda_{m'} > 0$*, then* $\frac{\lambda_m}{\Lambda_m} < \frac{\lambda_{m'}}{\Lambda_{m'}}$*.*

*Proof.* An active miner optimizes

$$\sum_{m=1}^{M}(\lambda_m + \lambda_{-m} + \Lambda_m)\left(1 - e^{-\rho R(1-f_m)\frac{\lambda_m}{\lambda_m+\lambda_{-m}+\Lambda_m}}\right) + (\lambda_A - \sum_{m=1}^{M}\lambda_m)\left(1 - e^{-\rho R}\right) \tag{12}$$

The marginal benefit of allocating hash rate to pool $m$ is

$$\begin{aligned} & 1 - e^{-\rho R(1-f_m)\frac{\lambda_m}{\lambda_m+\lambda_{-m}+\Lambda_m}}\left[1 - \rho R(1-f_m)\frac{\lambda_{-m}+\Lambda_m}{\lambda_m+\lambda_{-m}+\Lambda_m}\right] \\ = \quad & 1 - e^{-\rho R(1-f_m)y}\left[1 - \rho R(1-f_m)(1-y)\right], \end{aligned} \tag{13}$$

where we have used $\lambda_{-m} = (N-1)\lambda_m$ in equilibrium, $x = \frac{\lambda_m}{\Lambda_m}$, and $y = \frac{x}{Nx+1} < \frac{1}{N}$ which is increasing in $x$. Expression (13) is decreasing in $y$, and decreasing in $f_m$ as long as $\rho R(1-f)(1-y)y < 1$, which holds when $N > \rho R$. Therefore, if $\frac{\lambda_m}{\Lambda_m} > \frac{\lambda_{m'}}{\Lambda_{m'}} \geq 0$ and $f_m > f_{m'}$, (13) must be higher for pool $m'$, which implies the miner is better off allocating some marginal hash power from pool $m$ to pool $m'$ (which is feasible because $\lambda_m > 0$), contradicting the fact this is an equilibrium. If in addition $\lambda_{m'} > 0$, then $\frac{\lambda_m}{\Lambda_m} \geq \frac{\lambda_{m'}}{\Lambda_{m'}} \geq 0$ would also lead to a contradiction, yielding $\frac{\lambda_m}{\Lambda_m} < \frac{\lambda_{m'}}{\Lambda_{m'}}$.

Now among the group of pools charging the same fee, suppose the total allocation is $\hat{\lambda}_A$, then because (13) is strictly decreasing in $y$, we have $y$ being identical $\forall m$ in this group. Therefore,

$$\lambda_m = \frac{\hat{\lambda}_A}{\sum_{m' \in Group}\Lambda_{m'}}\Lambda_m. \tag{14}$$

for low enough $f$ and zero otherwise. □

The first statement in the proposition concerns a *Distribution Invariance in Equal-Fee Group*, which implies that without heterogeneous fees, we should not expect pool distribution to grow more dispersed or concentrated. But more importantly, the proposition reveals that if we view $\Lambda_m$ as the original size of the pool, then the pools' growth rates are inversely related to the fees they charge.

## 3.5 Empirical Evidence

The theoretical analyses in previous sections offer the following two testable predictions. Cross-sectionally, a pool with larger starting size tends to i) charge a higher fee, and ii) grow slower in percentage terms. We provide supporting evidence on these two predictions.

**Data description** Our data consist of two major parts: one on pool size evolution and the other on pool fee/reward type evolution. In part one, a pool's size (share of hash rates) is estimated from block relaying information recorded on the public blockchain (see BTC.com). Specifically, we count the number of blocks mined by a particular pool over some time interval, divide it by the total number of newly mined blocks globally over the same time interval; the ratio is the pool's estimated hash rate share. Balancing the trade-off between real-timeness and precision of estimation, we take the time interval to be weekly.[11]

In part two, the fee contract information is obtained from Bitcoin Wiki. We scrape the entire revision history of the website (477 revisions in total) and construct a panel of pool fee evolutions over time.[12] Pool fees are aggregated to quarterly frequency by simple average.

The two parts are then merged to construct a comprehensive panel data on pool size and fee evolution. Our main analysis will focus on the evolution of pool sizes at the quarterly frequency given potentially lagged adjustment. Table 1 in Section 2 provides summary statistics of the data.

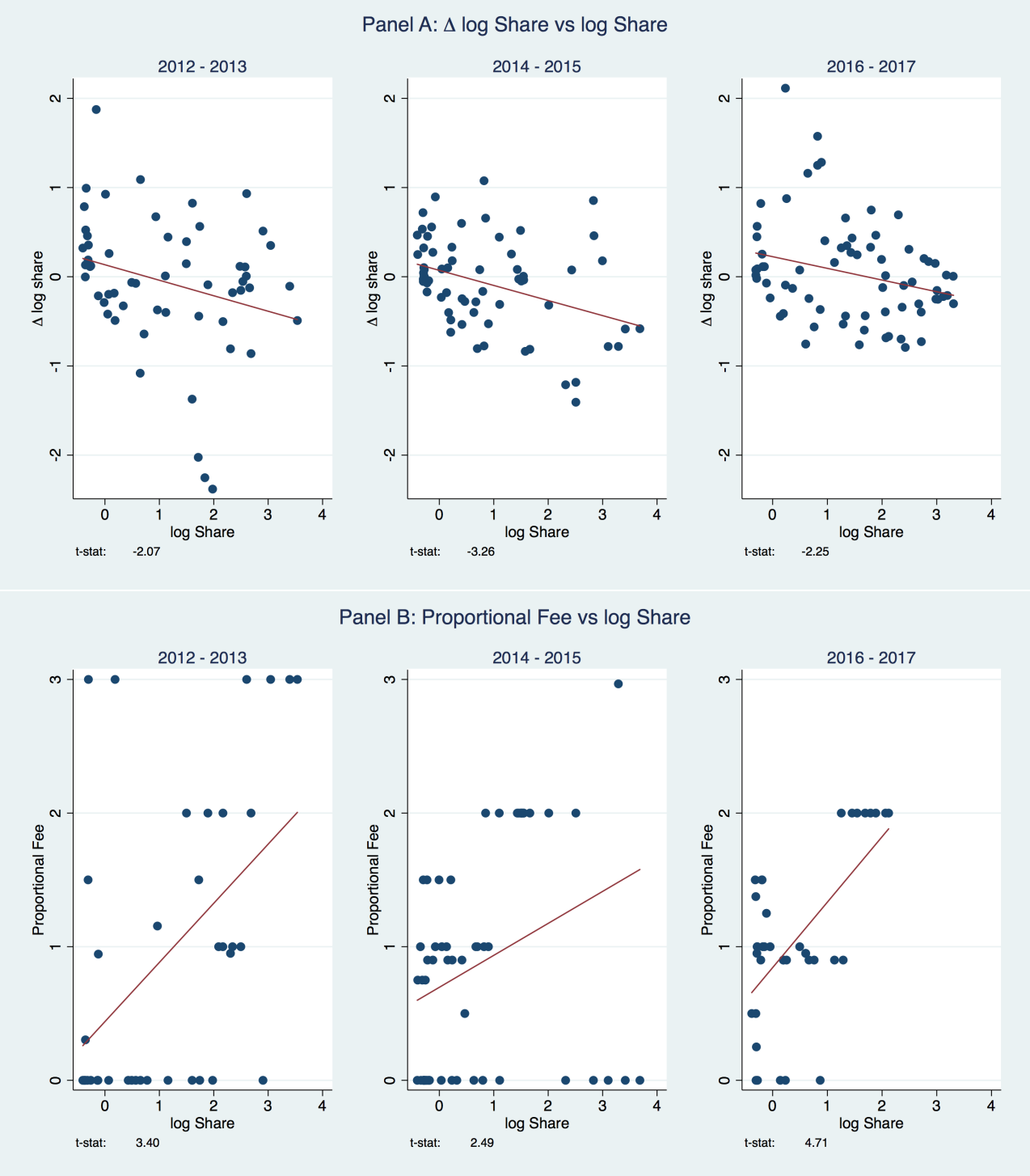
**Empirical results** Because our main predictions are concerning cross-sectional relationships, every quarter we first sort pools into deciles based on the start-of-quarter pool size (estimated hashrate share within the first week). We then treat each decile as one observation, and calculate the average proportional fee and average log growth rate across mining pools for each decile. Figure 3 shows the scatter plots for these decile-quarter observations, with Panel A (B) being the decreasing relationship between initial pool size and proportional fee (subsequent pool size growth rate). For robustness, we show the scatter plots for

[11] Our estimation procedure is standard. For example, blockchain.info provides real-time updates about estimated hashrate distribution over the past 24 hours, 48 hours, and 4 days using the same method. Bitcoinity tracks about 15 large mining pools' real time hashrate changes on an hourly basis. We favor weekly frequency over daily frequency because among all the pools that successfully find at least one block within a quarter, only (more than) 1.96% (42%) do not find any blocks within the first week (day) of that quarter. This is important because later analysis uses the estimated hash rate share within the first week as the initial pool size for the quarter.

[12] Two large pools are missing from the Wiki: Bixin (which was available in the wiki as HaoBTc prior to Dec 2016), and BTC.top, for which we fill their information through direct communication with the pools. Bitfury, which is also missing from the Wiki, is dropped as it is a private pool not applicable to our analysis.

Figure 3: **Pool Sizes, Fees, and Growths: Empirical Relationships**

This figure shows the binned plots of the changes in log Share (Panel A) and Proportional Fees (Panel B) against log Share. Share is the quarterly beginning (the first week) hash rate over total market hash rate. Fees are the quarterly averaged proportional fees. Within each quarter $t$, $\Delta logShare_{i,t+1}$, Proportional $\text{Fee}_{i,t}$, and $logShare_{i,t}$ are averaged within each $logShare_{i,t}$ decile, and these mean values are plotted for 2012-2013, 2014-2015, and 2016-2017, respectively. Red lines are the fitted OLS lines, with t-stat reported at the bottom. Data sources and descriptions are given in Section 3.5.



three two-year spans 2012-2013, 2014-2015, and 2016-2017, with t-statics for the regression coefficient reported underneath each plot.

As predicted by our theory, Figure 3 Panel A shows that larger pool grows in a slower pace, and Panel B shows that cross-sectionally a larger pool charges a higher fee. Importantly, all regression coefficients are statistically significant at 5% level for all three time periods.

# 4 Discussions and Extensions

In this section, we first take an economists' perspective on several important issues regarding (de)centralization within the burgeoning FinTech sector, and then consider an extension of our model with adjustable computational power $\lambda_A$ which is a relevant case in the long-run.

## 4.1 Centralization in Decentralized Systems

### The Blockchain Innovation

The emergence of FinTech and relatedly sharing economy is largely driven by the increasing preference for forming peer-to-peer connections that are instantaneous and open, which is transforming how people interact, work, and consume. Yet financial systems are often arranged around a series of centralized parties like banks and payments, clearing and settlement systems. Blockchain-based crypto-applications are part of an attempt to resolve the issue by creating the financial architecture for peer-to-peer transactions and reorganizing society into a series of decentralized networks.

The key innovation of the blockchain technology does not merely concern distributed ledgers or hash-linked data storage system. In fact, many technologies and applications preceding blockchain provide these functionalities already. It is the functionality of providing decentralized consensus that lies at the heart of the technology (e.g., Cong and He (2018)), and proof-of-work as manifested in Bitcoin mining plays an important role in the consensus generation process (e.g., Eyal (2015)). The decentralized consensus generation process arguably has two potential advantages. First, it avoids single points of failure; second, it reduces the market power of centralized third parties. Both advantages rely on the premise of decentralization. It has hence become a natural concern how centralized Bitcoin is.[13]

[13]See e.g. Gervais, Karame, Capkun, and Capkun (2014).

**Other Centralizing and Decentralizing Forces**

In this paper we have focused on the risk-sharing channel, which serves a centralizaing force, and the endogenous growth channel as a decentralizing force. There are many other channels that matter too. For example, Chapman, Garratt, Hendry, McCormack, and McMahon (2017), de Vilaca Burgos, de Oliveira Filho, Suares, and de Almeida (2017), and Cong and He (2018) discuss how the concern for information distribution naturally makes nodes in blockchain networks more concentrated.

Conventional wisdom in the Bitcoin community has proposed several reasons why a mining pool's size may be kept in check: 1) ideology: bitcoin miners, at least in the early days, typically have strong crypto-anarchism background, for whom centralization is against their ideology. While this argument may be true back in the early days of Bitcoin development, it has become a stretch today as Bitcoin develops into a hundred-billion-dollar industry; 2) sabotage: just like the single-point-of-failure problem in traditional centralized systems, large mining pools also attract sabotages (e.g. decentralized-denial-of-service (DDoS) attacks).Add DDoS papers. Indeed there are many self-reported DDoS attacks from peers against large mining pools (see Appendix **??** for a summary). While we believe sabotage to be an important force to control pool sizes, it is outside of the scope of this paper and left for future research; 3) trust crisis: it has been argued that Bitcoin's value builds on it being a decentralized system. Over-centralization by any single pool may lead to collapse in Bitcoin's value, which is not in the interest of the pool in question. In other words, a mining pool would voluntarily prevent over-concentration to avoid a self-hurting Bitcoin value crisis. Empirical evidence for this argument, however, is scarce. Indeed, there is no significant results when we associate the HHI of the mining industry with bitcoin prices, as well as investigate the bitcoin price responses around the GHash.io 51% attack in July, 2014.

## 4.2 Other Proof-of-Work Protocols

Our model can help us gain better understanding of the centralizing and decentralizing forces in blockchain-based systems beyond Bitcoin, especially for those that rely on proof-of-work. For example, Ethereum, a major blockchain-based platform with its native cryptocurrency having a market valuation second only to Bitcoin, also relies on a proof-of-work process. For each block of transactions, be it payments or smart contracting, miners use computation powers to solve for crypto-puzzles. More specifically, the miners run the block's unique header metadata through a hash function, only changing the 'nonce value',

which impacts the resulting hash value. If the miner finds a hash that matches the current target, the miner will be awarded ether and broadcast the block across the network for each node to validate and add to their own copy of the ledger. Again, the proof-of-work protocol (The specific proof-of-work algorithm that ethereum uses is called 'ethash') here makes it difficult for miners to cheat at this game, because the puzzles are hard to solve and the solutions are easy to verify. Similar to Bitcoin, the mining difficulty is readjusted automatically such that approximately every 12-15 seconds, a miner finds a block. Ethereum, along with other cryptocurrencies such as Bitcoin Cash (BCH), Litecoin (LTC), and ZCash (ZEC) that rely on PoW all witness pool formations.

## 4.3 Long-term Outcome with Adjustable $\lambda_A$

In our analysis thus far we have focused on the case where $\lambda_A$ is fixed, which leads to a multi-agent portfolio allocation problem with network externality. Fixed $\lambda_A$ is realistic in the short-run when miners cannot easily adjust the computation power they have. However, over the long-run, miners may acquire additional computation powers or sell mining chips to other miners. To analyze such "long-run" outcome, we consider adjustable $\lambda_A$.

To simplify our discussion, we make two additional assumptions: (1) each active miner is small and behave competitively towards a pool. (2) each pool is small and behave competitively towards the total computation power $\Lambda$ in the world. The first assumption is innocuous and natural, and resembles the "little k, Big K" situation in macro. The second is what we understand from forum discussions on setting pool fees. However, for very large pool owners to ignore the impact of fee change on the global computation power is rather heoric. What we derive in closed form below then only applies to a situation where we have many competitive small pools (which are still large relative to individual miners). The case with larger pools can be solved numerically later and all main results still hold.

The optimization problem for each active miner becomes

$$\max_{\lambda_m} \left[ -\rho\lambda_m c + \frac{(\lambda_m + \lambda_{-m} + \Lambda_m)\left(1 - e^{-\rho R(1-f_m)\frac{\lambda_m}{\lambda_m+\lambda_{-m}+\Lambda_m}}\right) + (\lambda_A - \sum_{m=1}^{M}\lambda_m)\left(1 - e^{-\rho R}\right)}{\Lambda D} \right] \tag{15}$$

for each $m$. Over the long-term, the active miner's problem is transformed from an allocation decision with constrained computation power to an endogenous computation-power

acquisition for each pool. This makes each pool's decision decoupled, at least for miners that are small and competitive.

Our earlier assumption that the miners take pool size as given is consistent with our earlier assumption $\rho R << N$, and allows us to expand the exponent up to second-order, to obtain the optimizer

$$\lambda_m^* = \frac{[R(1-f_m) - \Lambda Dc]\Lambda_m}{\rho R^2(1-f_m)^2 - NR(1-f_m) + N\Lambda Dc} \tag{16}$$

where we have used the fact that $\lambda_{-Am} = (N-1)\lambda_{Am}$ in equilibrium. It is important when we take this into consideration because when $\lambda_m$ or $f_m$ changes, the pool size also changes and need to be taken into consideration. It is obvious that $\lambda_{Am}$ is increasing in $\lambda_m$, but

$$\frac{\partial \lambda_m^*}{\partial f_m} \propto R(1-f_m) - 2\Lambda Dc \tag{17}$$

We note that in equilibrium this has to be non-positive, otherwise pool $m$ owner would always increase fee and gets more allocation. Alternatively, when $\Lambda$ is really big, increasing fee always decreases allocation.

Now consider pool owners' decision to set $f_m$s.

$$(N\lambda_m + \Lambda_m)[1 - e^{-\rho R f_m}] \tag{18}$$

One sufficient condition for the cross partial w.r.t. $f_m$ and $\Lambda_m$ to be positive if $R(1 - f_m) > 2\Lambda Dc$. As long as $R$ is large and the fee is not too high. Then by the results from monotone comparative statics (Milgrom and Shannon (1994), Topkis (1978)), the optimal $f_m$ is increasing in $\lambda_m$.

Therefore, even with adjustable hash rates, our earlier conclusions carry through when $R(1 - f_m) > 2\Lambda Dc$. To see if this condition holds currently, we note that 12.5 Bitcoins are rewarded every 10 minutes (on average). The total world wide hash rate is 4,547,580,033 GH/s, equivalent to 337,000 Antminer S9 mining rigs with an advertised hash rate of 13.5 TH/sec. The power needed is 1375 watts for six months for one Antiminer. In round numbers, that is 6000 kw hrs. The power costs is roughly \$400. Taking a fee of 5%, and a coin price $10,000$, we are with in the range where larger $Llambda_m$ leads to a larger $f_m$, approximately.

Now if we allow the pool owners to internalize the impact on $\Lambda$ when setting fees, we have to substitute $\Lambda$ in equation (16), we get a second order system of equations, which we

can numerically solve.

# 5 Conclusion

We view our paper's contribution to be three-fold. We first formally develop a theory of mining pools that highlights a natural centralizing force — risk-sharing, which is the main driver for creating mining pools in practice. We also document the industrial organization of the Bitcoin mining empirically, and provide evidence consistent with our theory. Third, we provide a new explanation for why Bitcoin mining maybe adequately decentralized over time. On the third one, while we do not claim that our explanation is necessarily the only or the best one, it does have a major advantage of being closely tied to the risk-sharing benefit that leads to the rise of mining pools in the first place. In this sense, it provides a backbone framework to analyze the interactions among mining pools, upon which other external forces (e.g. DDoS attacks) could be added on.

As a first paper to analyze the complicated mining pool dynamics, we have to leave many interesting topics to future research. For example, we do not take into account potential pool collusion or alternative pool objectives. Anecdotally, there is speculation that a large pool ViaBTC, along with allies AntPool and BTC.com pool, are behind the recent promotion of Bitcoin Cash, a competing cryptocurrency against Bitcoin. Hence these pools' behavior in Bitcoin mining may not necessarily be profit-maximizing. We also do not consider the effect of concentration in other stages along the vertical value chain of bitcoin mining (e.g. Bitmain, the owner of AntPool and BTC.com, as well partial owner of ViaBTC, is also the largest Bitcoin mining ASIC producer who currently controls 70% of world ASIC supply). We currently also do not look much into the entry/exit of mining pools and miners. Hence we caution readers to take our conclusions on the long-run industrial organization as a first-attempt benchmark result rather than a foregone conclusion.

# References

Beccuti, Juan, Christian Jaag, et al., 2017, The bitcoin mining game: On the optimality of honesty in proof-of-work consensus mechanism, Discussion paper, .

Berk, Jonathan B, Richard Stanton, and Josef Zechner, 2010, Human capital, bankruptcy, and capital structure, *The Journal of Finance* 65, 891–926.

Biais, Bruno, Christophe Bisiere, Matthieu Bouvard, and Catherine Casamatta, 2018, The blockchain folk theorem, .

Burdzy, Krzysztof, David M Frankel, and Ady Pauzner, 2001, Fast equilibrium selection by rational players living in a changing world, *Econometrica* 69, 163–189.

Calvo, Guillermo A, 1983, Staggered prices in a utility-maximizing framework, *Journal of monetary Economics* 12, 383–398.

Chapman, James, Rodney Garratt, Scott Hendry, Andrew McCormack, and Wade McMahon, 2017, Project jasper: Are distributed wholesale payment systems feasible yet?, *Financial System* p. 59.

Cong, Lin William, and Zhiguo He, 2018, Blockchain disruption and smart contracts, .

Cong, Lin William, Ye Li, and Neng Wang, 2018, Tokenomics: Dynamic adoption and valuation, *BFI Working Paper*.

de Vilaca Burgos, Aldenio, Jose Deodoro de Oliveira Filho, Marcus Vinicius Cursino Suares, and Rafael Sarres de Almeida, 2017, Distributed ledger technical research in central bank of brazil, .

Dimitri, Nicola, 2017, Bitcoin mining as a contest, *Ledger* 2, 31–37.

Easley, David, Maureen O'Hara, and Soumya Basu, 2017, From mining to markets: The evolution of bitcoin transaction fees, .

Economist, The, 2017a, Learning the lessons of equihack, *The Economist September 16, 2017* September 16, 14.

——— , 2017b, A yen for plastic, *The Economist* Nov 4, 2017, 72.

Eyal, Ittay, 2015, The miner's dilemma, in *Security and Privacy (SP), 2015 IEEE Symposium on* pp. 89–103. IEEE.

——— , and Emin Gün Sirer, 2014, Majority is not enough: Bitcoin mining is vulnerable, in *International Conference on Financial Cryptography and Data Security* pp. 436–454. Springer.

Fisch, Ben, Rafael Pass, and Abhi Shelat, 2017, Socially optimal mining pools, in *International Conference on Web and Internet Economics* pp. 205–218. Springer.

Gencer, Adem Efe, Soumya Basu, Ittay Eyal, Robbert van Renesse, and Emin Gün Sirer, 2018, Decentralization in bitcoin and ethereum networks, *arXiv preprint arXiv:1801.03998*.

Gervais, Arthur, Ghassan Karame, Srdjan Capkun, and Vedran Capkun, 2014, Is bitcoin a decentralized currency?, *IEEE security & privacy* 12, 54–60.

Harris, Milton, and Bengt Holmstrom, 1982, A theory of wage dynamics, *The Review of Economic Studies* 49, 315–333.

Harvey, Campbell R, 2016, Cryptofinance, *Working Paper*.

Hayek, Friedrich August, 1945, The use of knowledge in society, *The American economic review* 35, 519–530.

He, Zhiguo, and Wei Xiong, 2012, Dynamic debt runs, *Review of Financial Studies* 25, 1799–1843.

Holmström, Bengt, 1982, Moral hazard in teams, *The Bell Journal of Economics* pp. 324–340.

Huberman, Gur, Jacob D Leshno, and Ciamac C Moallemi, 2017, Monopoly without a monopolist: An economic analysis of the bitcoin payment system, .

Khapko, Mariana, and Marius Zoican, 2017, 'smart' settlement, *Working Paper*.

Kiayias, Aggelos, Elias Koutsoupias, Maria Kyropoulou, and Yiannis Tselekounis, 2016, Blockchain mining games, in *Proceedings of the 2016 ACM Conference on Economics and Computation* pp. 365–382. ACM.

Kroll, Joshua A, Ian C Davey, and Edward W Felten, 2013, The economics of bitcoin mining, or bitcoin in the presence of adversaries, in *Proceedings of WEIS* vol. 2013. Citeseer.

Li, Jiasun, 2015, Profit-sharing, wisdom of the crowd, and theory of the firm, *Discussion Paper*.

Malinova, Katya, and Andreas Park, 2016, Market design for trading with blockchain technology, *Available at SSRN*.

Milgrom, Paul, and Chris Shannon, 1994, Monotone comparative statics, *Econometrica: Journal of the Econometric Society* pp. 157–180.

Nakamoto, Satoshi, 2008, Bitcoin: A peer-to-peer electronic cash system, .

Nanda, Ramana, Robert F. White, and Alexey Tuzikov., 2017, Blockchain, cryptocurrencies and digital assets, *Harvard Business School Technical Note* pp. 818–066.

Nayak, Kartik, Srijan Kumar, Andrew Miller, and Elaine Shi, 2016, Stubborn mining: Generalizing selfish mining and combining with an eclipse attack, in *Security and Privacy (EuroS&P), 2016 IEEE European Symposium on* pp. 305–320. IEEE.

Raskin, Max, and David Yermack, 2016, Digital currencies, decentralized ledgers, and the future of central banking, Discussion paper, National Bureau of Economic Research.

Rosenfeld, Meni, 2011, Analysis of bitcoin pooled mining reward systems, *arXiv preprint arXiv:1112.4980*.

Saleh, Fahad, 2017, Blockchain without waste: Proof-of-stake, .

Sapirshtein, Ayelet, Yonatan Sompolinsky, and Aviv Zohar, 2015, Optimal selfish mining strategies in bitcoin, *arXiv preprint arXiv:1507.06183*.

Schrijvers, Okke, Joseph Bonneau, Dan Boneh, and Tim Roughgarden, 2016, Incentive compatibility of bitcoin mining pool reward functions, in *International Conference on Financial Cryptography and Data Security* pp. 477–498. Springer.

Stiglitz, Joseph E, 1974, Incentives and risk sharing in sharecropping, *The Review of Economic Studies* 41, 219–255.

Topkis, Donald M., 1978, Minimizing a submodular function on a lattice, *Operations Research* 26(2), 305–21.

Weiss, Mitchell, and Elena Corsi, 2017, Bitfury: Blockchain for government, *HBS Case Study* January 12, 818–031.

Wilson, Robert, 1968, The theory of syndicates, *Econometrica* pp. 119–132.

Yermack, David, 2017, Corporate governance and blockchains, *Review of Finance* p. rfw074.