Token-based Decentralized Governance, Data Economy and Platform Business Model

March 15, 2024

ABSTRACT

The growing importance of data as an input for platform businesses, combined with users' concerns about sharing their personal data, raises questions how to balance efficiency and fairness in governing these platforms. As a potential solution, we examine the decentralized governance of a platform that utilizes governance tokens issued to users. We explore the incentives for founders to establish platforms with decentralized governance and the resulting benefits for users. Decentralized platforms lead to a greater surplus for users compared to traditional, centrally governed platforms, enhancing fairness. In 'data-heavy' platforms, a buy-back market for governance tokens emerges, enabling the founder to promote early platform adoption by committing to future transferability of tokens among users. Consequently, the platform's founder can achieve equal or greater output compared to a centrally governed platform, which provides a rationale for token issuance. Regulations that limit data sharing negatively impact stakeholders of decentralized platforms, while they might protect users in centralized ones. Token-based decentralized governance offers a way to align the interests of platform founders with those of users.

JEL Classifications: D21, D23, D24 G34, G38, O33

Keywords: Decentralized governance, Governance token, Tokenomics, DAO, Data economy, Personal data, Privacy concerns, Platform business model

I. Introduction

Firms operating on the platform business model are increasingly supported by global digital technology infrastructures. The personal data collected from users through their participation on platforms, combined with advancements in data processing technologies like machine learning and artificial intelligence, have led to the emergence of data as a critical input to production used throughout the economy.[1] Technology companies that operate the largest platforms, for example, act as owners of the personal data they have collected and, for this reason, are under scrutiny by the general public and policymakers due to their market power, handling of consumer privacy, and ethical standards in the collection and utilization of personal data.[2]

Despite the important role that personal data plays in the economy, there are no well-established mechanisms that would govern the use of personal data in production.[3] The assignment of property rights to data plays a pivotal role in shaping its utilization, thereby influencing output, private value creation, and social welfare. The emergence of the data economy built on the platform business model thus raises pressing questions about how to govern the use of personal data in a way that creates a fair and efficient digital future for our society.

Balancing user privacy and economic efficiency, Jones and Tonetti (2020) suggests allocating data property rights to consumers. This approach encourages consumers to carefully weigh their concerns regarding privacy against the economic gains derived from data sharing and thus use in

[1]Recently, scholars have increasingly focused on studying the role of data as an input in the production process, so called 'Data Economy'. Unlike traditional inputs such as labor or real capital, data is generated through agents' economic activities and is non-rival, that is, data can be used simultaneously by many firms. Veldkamp and Chung (2023) offers an overview of how the growth of data in the economy interacts with key issues of macroeconomic policy and economic growth. Farboodi and Veldkamp (2022) provide a valuation model of transaction-generated data and examine the aggregate economic consequences of data accumulation. Jones and Tonetti (2020) and Cong, Wei, Xie, and Zhang (2022b) stress the non-rivalry nature of data when used as input to production. Bessen (2022) highlight the complementarity between software and data.

[2]The world's top companies—Apple, Google, Amazon, Microsoft, Facebook, and Tencent, among them—are based on platform business models. These firms work to build a diverse ecosystem that not only offers Internet search, social media, and e-commerce, but also provides an infrastructure on which new platforms can be built. U.S. Congress held a series of hearings on issues stemming from the onset of platform business models, for example, platforms and Market Power, Part 3: The Role of Data and Privacy in Competition, platforms and Market Power, Part 6: Examining the Dominance of Amazon, Apple, Facebook, and Google, or Protecting Consumer Privacy in the Era of Big Data. See also, The Ethical Dilemma at the Heart of Big Tech Companies or Views of Big Tech Worsen; Public Wants More Regulation.

[3]Ramirez, Brill, Ohlhausen, Wright, and McSweeny (2014) describe the practices of data brokers, discuss costs and benefits of transacting personal data, and highlight non-transparency in handling personal data. Galperti and Perego (2022) show that opening private markets for data does not necessarily lead to efficiency even if markets are perfectly competitive. Cong and Mayer (2023) study inefficiencies in data collection and show that these inefficiencies cannot be fully addressed by data sharing proposals, user privacy protections, platform commitments, and markets for data.

production. Even if consumers hold property rights to data, they may still find themselves in a disadvantaged position relative to firms due to weak bargaining power and inadequate information. This scenario results in an equilibrium that is characterized by inefficiencies in terms of both output and societal welfare (Cong, Xie, and Zhang, 2021b; Cong et al., 2022b; Ichihashi, 2021). To address these inefficiencies, prior work has explored potential solutions, such as user unions (Cong and Mayer, 2023) or contract markets that enable platforms and consumers to specify the terms of data utilization (Galperti and Perego, 2022). Overall, this literature underscores the absence of effective solutions for the issue of data governance.

In this paper, we examine decentralized governance mechanism supported by blockchain technology as a potential solution to these challenges. This blockchain-based governance mechanism allows users to directly participate in the platform's governance through the use of governance tokens, effectively giving them a say in how their data is collected and utilized. In this context, several important questions emerge, mainly: Do platform founders want to adopt such governance mechanism, and how does this mechanism impact the surplus of platform users? We aim to provide answers to these questions.

The blockchain-based governance structure of a platform is typically established as a Decentralized Autonomous Organization (DAO).[4] Programmed rules of the DAO determine what changes to the platform can be made and what consensus is needed to make these changes. Users participate in platform governance by holding a governance token which gives them the right to vote on proposals regarding platform updates.[5] In DAOs, the cost of collective decision-making is lowered by the decentralized blockchain technology which ensures frictionless voting and automatic full

[4]An example of a DAO is Compound, a decentralized finance (DeFi) platform where users can lend and borrow cryptocurrencies. Users can participate in Compound's governance by voting about proposals to update the platform, such as changes in the Compound protocol, fee and parameter adjustments, or opening markets in new cryptocurrencies, using a governance token COMP. One COMP represents one vote. Apart from the initial distribution of COMP tokens, users also receive COMP as a reward for their economic activities on the platform. Other examples of governance tokens are UNI in platform Uniswap and MKR in platform MakerDAO. Unlike the governance of traditional firms, decentralized platform governance is flexible and less regulated. Platform founders design properties of governance tokens and governance processes at their will when the platform is set up. Platform founders' choices include but are not limited to, considering how to distribute tokens, how to design the voting mechanism, whether to allow the platform to buy back the governance tokens, and how to set the timing and price of such buy-backs. Empirically, we observe a wide range of governance choices: Compound and Uniswap have different voting thresholds for a proposal to be passed; Maker allows buy-back of their token, while Compound does not.

[5]Unlike utility tokens, which are usually considered as means of payment or 'ticket' to service provided by a platform, governance token is associated with rights such as voting and, sometimes, cash flow rights. Governance tokens are typically distributed to platform users for free as a 'reward' for using the platform, and are thus not used to raise capital through initial coin offerings (ICOs).

enforcement of the voting outcome.

General public has divergent views on the merit of such governance structures and stresses the ultimate control of investors over business organizations.[6] However, as highlighted in a seminal work by Hansmann (2000): *"...while we tend to take for granted that business enterprise is organized in the form of investor-owned firms, investor ownership is not a logically necessary concomitant of free markets and free enterprise."* Furthermore, Hansmann argues that a form of business organization is endogenously determined and depends on technologies the society has available at any given point in time.

We first present a simplified illustrative example that synthesizes the concepts of the data economy, platform business model, and decentralized governance mechanism. Within the healthcare domain, there exists a specialized field concerning the judicious utilization of sensitive data. In this context, previous literature focuses on studying patient privacy concerns and data-enriched technology adoption (Miller and Tucker, 2009, 2018). Setting our example in this context, we consider a healthcare platform that generates medical services by leveraging users' personal data, such as biometric/genetic information. Users share their personal data with the platform. In return for contributing their data, users receive governance tokens as rewards. The platform employs software, e.g., based on artificial intelligence, to create medical services, e.g., recommendations, which are the platform's outputs. To influence the governance of the platform, users utilize their governance tokens to cast votes on proposals. These proposals may pertain to matters such as expanding the scope of data sharing, a move that could enhance the quality of the platform's services but may also elevate concerns related to data privacy. The platform's governance rules are set by platform founders by pre-programmed codes at the time the platform is set up. Platform founder's payoff depends on the volume of transactions on the platform.

We introduce a model to assess the efficiency of decentralized governance from the platform founder's perspective, as well as the level of fairness that users can achieve through their partici-

[6] *"You don't own 'web3'. The VCs [venture capitalists] and their LPs [limited partners] do. It will never escape their incentives. It's ultimately a centralized entity with a different label. Know what you're getting into..."* — Jack Dorsey, co-founder of Twitter and crypto fan, tweets a scathing criticism of the idea that "Web3" will democratize the Internet. Relatedly, Elon Musk cheekily tweeted, *"Has anyone seen 'web3'? I can't find it"* – to which Dorsey replied, *"It's somewhere between a and z,"* a dig at the venture capital firm Andreessen Horowitz, known as a16z, which is one of the largest investors in startups that rely on the decentralized blockchain technology. Tim Berners-Lee, the computer scientist credited with inventing the World Wide Web, dismisses 'Web3' and presented his own proposal for reshaping the Internet, called "Web 3.0" that also aims to give users control of their own data, including how it's accessed and stored.

pation in the platform's governance. The platform relies on users' personal data to produce goods for consumption. The platform's output increases with the extent of data that users share with the platform, and the extent of platform adoption by users. The output of the platform is fully consumed by its users. A key assumption in our model is that by joining the platform, heterogeneous users share their personal data with the platform and, by doing so, incur a non-pecuniary cost that depends on the extent of data shared. Following previous literature (Goldfarb and Tucker, 2012; Lin, 2022), we assume that users have heterogeneous costs, as they may have, for example, different concerns about the privacy of their personal data. Given that a user's cost and the platform's output increase as the extent of data shared increases, the key trade-off in our model is between the extent of data shared and the extent of platform adoption by users.

The objective of the platform founder is to maximize the total output of the platform by devising the properties of the platform's governance token.[7] The governance tokens are distributed for free as rewards to users who engage in the platform's production activities by sharing their data. Additionally, governance tokens function as representations of users' voting rights in the platform's governance processes. The platform founder's decision consists of setting a voting rule based on distributed tokens, as well as deciding whether to establish a market for the platform to buy back the governance token from token holders at a 'buy-back' price. To distinguish between the two governance regimes, we refer to the one without a buy-back mechanism as the "*voting-only regime*" and the other one, which includes both buy-back and voting mechanisms, as the "*buy-back and voting regime*".

The model has two periods and three dates. An economy consists of a platform capable of production, potential users of the platform (consumers of the outputs), and a platform founder. In the beginning, the founder sets up the platform with the minimum (initial) extent of data sharing that allows it to start producing. Governance tokens are allocated to users as a by-product of users' participation in the first-period production by the platform. Subsequently, after the end of the first period, platform users make a key decision as they vote on the increment in the extent of data sharing the platform will require from its users. Effectively, the users are deciding about the platform's growth. The voting rule requires a fraction of users to agree for the voting result to be passed and implemented. In the real world, governance tokens can be issued by a platform to raise

[7]Typically, the payoff of a platform founder increases with the volume of transactions conducted on the platform.

capital while retaining voting rights at a later stage. We concentrate solely on the governance aspect of token issuance in order to demonstrate that, even without the financing function, governance tokens are valued by platform stakeholders as they help address coordination challenges among platform's stakeholders.[8]

We characterize the solution to the platform founder's problem in *voting-only regime* and *buy-back and voting regime*. We then compare this optimal decentralized platform solution to two benchmarks. In the first benchmark, the platform is operated by an owner who maximizes the platform's output by directly deciding the increment in the extent of data sharing and by using transfers among users in order to entice them to join the platform in both periods. This *centrally governed platform* benchmark resembles a firm with traditional corporate governance. In the second benchmark, the platform is operated by a *social planner* who maximizes the total surplus of the platform's users by directly deciding which users participate in the platform and the increment in the extent of data sharing.

An important property of our model is that both, platform user adoption and the increase in the extent of data sharing required by the platform, which are jointly affecting the platform's output, are endogenously determined. The decision to increase the extent of data sharing is determined by the outcome of the voting process among token holders. Anticipation of more data sharing enhances the platform's productivity per user while, at the same time, also increases users' costs from participating in the platform. Considering this trade-off, users make decisions regarding their participation in the platform at the beginning of each period, as well as whether to leave the platform through token buy-back, which in turn influences user adoption.

The governance rules implemented by the founder at the setup of a platform influence the platform's user adoption and the extent of data sharing in the following way. Firstly, all else being equal, a decrease in the voting rule threshold restricts high-cost users from expressing their preferences in voting, which leads to a larger increase in data sharing. A larger increase in data sharing simultaneously sets a high bar for users to join the platform while also increasing their gains from production in the later period. Secondly, the possibility for users to leave the platform through token buy-back affects the composition of governance token holders and influences initial

[8]Bakos and Halaburda (2022), Li and Mann (2018), and Sockin and Xiong (2023a) argue that a significant advantage of decentralized token issuance is the reduction of potential commitment issues, alleviation of underinvestment problems, and the facilitation of the platform's adoption of new marketplaces.

user adoption. The effect of a change in the buy-back price on user adoption in the latter period and the magnitude of the increase in data sharing is mixed. On the one hand, a higher buy-back price attracts more users to join the platform early, resulting in a voting group that may prefer a smaller increase in data sharing. On the other hand, a higher buy-back price induces high-cost users to sell their governance token, leading to a voting group that tends to support a higher increase in data sharing. The overall effect depends on which of the two forces dominates.

We classify platforms into 'data heavy' and 'data-light', depending on the minimum extent of data sharing required for the platform to start producing. The distinction between the two types of platforms is determined by a threshold that depends only on the legal restriction on the extent of data-sharing in the economy, which we take as exogenous. The properties of the optimal decentralized governance solution depend on the platform type, and our key results concern 'data-heavy' platforms.

Our model provides two primary insights. First, we show that in the voting-only regime and both types of platforms, the founder can always achieve a platform output that is weakly higher compared to the social planner solution. At the same time, the output is always weakly lower compared to the centrally governed platform. The reverse ranking is obtained for the total user surplus generated by the platform. The voting-only decentralized governance mechanism thus stands "in the middle" in terms of the trade-off between user surplus and platform output, as compared to the two benchmark governance regimes we consider.

Our second key result is that in 'data-heavy' platforms, the buy-back market for governance tokens exists and tokens are traded at a positive price, while the buy-back market does not arise in 'data-light' platforms. The positive price stems from the retrading of tokens which provides users with voting rights and tickets to the second-period service provided by the platform. The sources of value of governance token are similar to what is mentioned in Brunnermeier, Merkel, and Sannikov (2022). Importantly, using the buy-back mechanism, the platform founder can achieve an equal or higher platform output compared to the centrally governed platform. This finding highlights why—in case of 'data-heavy' platforms—founders may prefer to set them up with decentralized token-based governance rather than as firms with traditional corporate governance structure. This result also underscores the advantages of issuing tradable governance tokens and distributing them to users at no cost. These two main results hold when data is at least as important as real capital in the platform's production function.

When data and real capital have equal importance in the platform's production function, that is, they have the same elasticity with respect to the platform's output, the founder can achieve a weakly higher total output by using a *buy-back and voting regime* compared to a *voting-only regime*. By allowing token buy-back, the platform's founder can commit to a mechanism that induces early adoption of the platform by allowing transfers among users at a later date. This result further suggests that issuing tradable tokens is an important element of optimal decentralized governance, and highlights the significance of token issuance in cases where early adoption is crucial for the platform's success.

The optimal decentralized governance rules in 'data-heavy' platforms—the optimal voting threshold and the optimal buy-back price—generally follow an inverse U-shape pattern as the initial data requirement increases. Furthermore, for intermediate levels of the initial data requirement, where the tradeoff between the extent of data sharing and user adoption is severe, a small change in the initial data requirement level results in marked changes in the optimal voting threshold and token buy-back price. In other words, for intermediate levels of the initial data requirement, small changes in the platform type have a significant impact on the optimal governance rules.

Finally, we discuss the implications of a government policy intervention that imposes restrictions on the extent of user data sharing in the economy. In a centrally governed platform, such regulation is effective as it prevents the owner of the platform from draining user surplus. However, these restrictions can lead to a reduction in total output and user surplus in platforms with optimally designed decentralized governance.

To summarize, we emphasize the benefits of token-based decentralized governance for managing digital ecosystems that heavily rely on personal data as inputs. Our model suggests that token-based decentralized governance in the platform economy has some potential to align the incentives of platform founders with the interests of its users compared to traditional firms. We show why platform founders may find it optimal to freely distribute governance tokens to platform users and why such tokens have a non-zero price in equilibrium. We also demonstrate the challenge faced by platform founders in identifying optimal decentralized governance schemes due to the complexity of the tradeoffs involved.

We build on the literature that asks how to organize economic activity. In Alchian and Demsetz

(1972), a firm is a team with joint complementary production. In this tradition, we study a DAO in the context of the production economy where data is a crucial input. Our setup of a platform with decentralized governance resembles customer cooperative (i.e., a customer-owned enterprise) that were prominent in housing (see Hansmann (2000), Ch. 11). Two features that Hansmann considers to be the main drivers of housing cooperatives—the homogeneity of users of housing and the cost of collective decision making—are also important in our decentralized platform setting. First, potential users self-select into joining the platform which makes the platform's user base more homogeneous in terms of costs from sharing data with the platform. Second, in our setting, the cost of collective decision-making is lowered by the decentralized blockchain technology which ensures a frictionless voting system and automatic enforcement of the voting outcome. This effectively means that, by devising governance rules at the time the platform is set up, the platform's founder can commit the platform to future outcomes that depend on users' votes. The transparency of the blockchain also supports our assumption that once set, the platform's governance rules are common knowledge. The adoption of this technological innovation is one of the key forces driving the popularity of decentralized governance.

Corporate governance literature focuses on investor-owned firms. The findings of this literature thus cannot be directly applied to discuss the efficiency and fairness of DAOs with governance tokens. First, differently from equity and debt, governance tokens are typically given as a reward for conducting economic activity on platforms and users receive them without any pecuniary payment. Second, compared to a centralized firm, where the CEO or the board of directors make most decisions, all users possessing a governance token can directly participate in platform governance. While shareholder proposals in typical investor-owned firms are not binding, platform rules execute automatically if approved through voting. Third, while board members in a corporate setting are predetermined when they vote, properties of a governance token influence users' incentive to participate in voting. From this literature, our paper is closely related to Broccardo, Hart, and Zingales (2020) who study social welfare in the context of a traditional investor-owned firm where some agents' preferences depend on the social impact of their decisions. We differ by analyzing the social welfare of an enterprise that is owned by self-interested consumers.

The paper is organized as follows. Section II presents the literature review on data economy and DAO governance. Section III describes the four settings we examine: *voting-only* decentralized governance regime, *buy-back and voting* decentralized governance regime, *centrally governed platform* benchmark, and *social planner* benchmark. Section IV presents the solution of the *voting-only* regime and *buy-back and voting* regime. In Section V, we compare decentralized platform governance with the two benchmarks. Section VI concludes.

II. Literature review

A. Data economy

A.1. Data and data economy

In Romer (1990), data refers to information that is not inherently instructional for creating a product but can still be valuable in the production process. This includes biographical data, personal information, and transaction data. A similar concept analogous to data is the notion of an "idea." An idea refers to information that offers instructions for generating outputs. In simpler terms, an idea represents a production function, while data serves as a factor of production.

As a crucial input in the production function, some researchers focus on differentiating data from other inputs like capital and labor. On one hand, Arrieta-Ibarra, Goff, Jiménez-Hernández, Lanier, and Weyl (2018) contend that it might be more appropriate to treat data as labor, rather than capital. This perspective considers individual ownership, incentives for ordinary contributions to production, and the countervailing power to create a data labor market. On the other hand, Cong et al. (2021b) posit that the characteristics of endogenous generation, dynamic nonrivalry, and flexible ownership of data set it apart from both labor and capital.

The field of data economics explores the relationship between data and the overall economy in various ways, including GDP measurement, monetary neutrality, growth, and firm dynamics (Veldkamp and Chung, 2023). Among these fields, researchers are particularly interested in assessing the level and growth impact of data on the economy. The level effects refer to the nonrivalrous nature of data. Nonrivalry leads to increasing returns, implying that there can be societal benefits when data is widely utilized across firms, even when privacy considerations are taken into account (Jones and Tonetti, 2020). Conversely, growth effects focus on the idea that data can be endogenously generated as a result of economic activity, accumulate over time, and contribute to long-run economic growth (Cong et al., 2021b; Cong and Mayer, 2023). In this paper, for simplicity, we do not consider the endogenous generation of data. Instead, we treat data as an endowment for users in the economy. Users can share their data for production, and contributing data is associated with privacy costs. In this sense, data shares some similarities with labor, where users contribute their labor to production and sacrifice their leisure time.

Data property rights can be attributed to firms, consumers, and social planners. In Jones and Tonetti (2020), firms tend to hoard data to maintain their competitive advantage, while consumers weigh their concerns for privacy against the economic gains derived from sharing data more broadly. Consumers are generally inclined to share their data. Our paper closely aligns with the key conclusion in Jones and Tonetti (2020), which suggests that assigning data property rights to consumers can lead to allocations that are close to the social planner's solution. Similar results are found in Miller and Tucker (2018), where they explore three types of policies and their impact on individual data-sharing behavior. The first policy involves sending users notifications about potential privacy risks before they share data, the second focuses on restricting discriminatory usage of data by firms, and the third grants control rights to individuals over data redisclosure. Their findings indicate that only the third policy encourages data sharing, while the first policy deters individuals from supplying the data.

However, even when consumers or users are granted data property rights, they may still find themselves in a relatively weak position, resulting in an economy that is inefficient in terms of both total output and overall welfare. Cong et al. (2021b) introduces a model where consumers sell their data to innovation intermediate firms for research purposes. The findings suggest that this type of economy tends to generate lower social welfare, primarily because the R&D sector underemploys labor while overutilizing data. Additionally, in Cong et al. (2022b), they study endogenous data usage across the innovation and production sectors. It is revealed that consumers' inability to fully internalize knowledge spillover costs, combined with firms' market power, leads to data being undervalued and an inefficient data supply. Ichihashi (2021) investigates the dynamics of competing intermediaries purchasing data from users. Considering multiple sharing possibilities, these intermediaries often offer low compensation. Intermediaries together with downstream firms drain surplus from consumers.

Researchers have proposed several potential solutions to address these inefficiencies. In Cong and Mayer (2023), it is observed that when users are separated, they often fail to fully grasp the impact of their data contributions on factors such as service quality for other users, market concentration, and the incentives of platforms to invest in data infrastructure. This can result in the inefficient over- or under-collection of data. One solution suggested is the formation of user unions, where cooperative users can work together to address antitrust concerns and enhance consumer protection. In Galperti and Perego (2022), they find that granting consumers control over their data can lead to lower social welfare and, at best, achieve second-best efficiency. To mitigate this, a possible solution is to open a large set of markets, enabling platforms and consumers to enter into contracts that specify how their data will be used.

In our paper, we present a novel solution: a decentralized platform with a governance token. Previous research has highlighted two key features for enhancing welfare and safeguarding consumer interests: cooperative consumers and granting consumers the right to monitor how their data is utilized. These attributes are seamlessly integrated into a decentralized platform. With the inclusion of a governance token, users become owners of the platform, exerting full control over the usage of their data. Simultaneously, they collectively participate in decision-making through voting to determine how their data should be utilized.

A.2. Privacy concern and data sharing

Acquisti, Taylor, and Wagman (2016) explore the complex intersection of privacy and data sharing in its literature review. Privacy, being a multifaceted concept, poses challenges in its precise definition as it holds varying interpretations among individuals. Nevertheless, a common thread unifying these interpretations is that, privacy pertains to the demarcation between the self and others, delineating the boundaries between the private and the public. It is essential to recognize that privacy should not be perceived as the antithesis of sharing; rather, it encompasses the ability to exercise control over the act of sharing.

Prior work documented evolution of users' privacy concerns over time and their heterogeneity among different user groups. Using survey data, Goldfarb and Tucker (2012) find that users' privacy concerns have risen over time, and users' privacy concerns vary with age. In Lin (2022), the intrinsic value of privacy is highly heterogeneous across consumers and categories of data. Several studies employ structural models to assess privacy concerns (Tang, 2019; Lin, 2022; Quan, 2022). Abowd and Schmutte (2019) introduce an algorithm from the computer science sector to quantify the value of privacy and the advantages derived from data sharing. Kummer and Schulte (2019) examine a market-driven trade-off between monetary considerations and privacy.

From the perspective of both social planners and individuals, an optimal approach to balance data sharing and privacy involves operating at the point that the marginal cost of increasing privacy equals the marginal benefit of disclosing data. However, in data-heavy industries, such as healthcare and advertising, the empirical literature on privacy regulation has found that policies aim to protect users' privacy hamper the dissemination and application of data-enriched technologies as well as the consumer's welfare (Miller and Tucker, 2009, 2018; Goldfarb and Tucker, 2011; Quan, 2022).

From a user's perspective in digital economies, consumers frequently face significant challenges in making well-informed decisions regarding their privacy. This is due to their weak position characterized by incomplete or asymmetric information concerning the timing of data collection, the intended purposes for collection, and the resulting ramifications from collection (Acquisti et al., 2016). On the other hand, from firms' perspective, Fainmesser, Galeotti, and Momot (2023) investigates the incentive for digital business platforms to collect and safeguard user data. This research reveals that a business model oriented towards data-driven revenue tends to accumulate more user data and provide a higher degree of data protection compared to businesses primarily driven by usage. However, it also demonstrates that businesses may exhibit tendencies to either over-collect or under-collect user data and may provide excessive or inadequate levels of protection, depending on the specific social context in which they operate.

B. DAO governance, tokens, and voting

B.1. DAO governance

DAOs, or Decentralized Autonomous Organizations, are entities designed to facilitate coordination among a large number of individuals. These organizations can have objectives that are either for-profit or nonprofit, with many of them focusing on activities related to cryptocurrencies or Web3 (Appel and Grennan, 2023a). For instance, DAOs enable customers to directly connect with servers, allowing for the maximization of both consumer and server profits through full disintermediation Cachon, Dizdarer, and Tsoukalas (2022). However, the potential fragmentation of the network itself may hinder coordination among users. To address this challenge, Arruñada and Garicano (2018) suggest the development of new forms of ”soft” decentralized governance, including anarchic, aristocratic, democratic, and autocratic approaches.

Tokens are a category of cryptocurrencies closely associated with DAOs. Typically, we discuss two main types of tokens: utility tokens and governance tokens. Utility tokens, often referred to as stablecoins, serve as tickets for accessing services on a platform. They are characterized by high liquidity, transparent market, and volatile prices. On the other hand, governance tokens, which are typically linked to voting rights, are traded less frequently, and their markets are opaque.

The relationship between DAO governance and tokens is intricate. On one hand, governance has an impact on token prices. In Appel and Grennan (2023b), it was found that governance that encourages broad participation in decision-making or enhances security is linked to positive abnormal returns for tokens, while the failed improvement proposals are associated with negative abnormal returns. On the other hand, the design of the token structure affects the efficiency of the DAO. Mayer (2019) suggests that a dual token structure comprising a governance token and a stablecoin can mitigate the crowding-out effect by speculators and promote adoption of the platform.

Our paper places a primary emphasis on governance tokens and investigates how the design of both the platform and the token can influence overall efficiency and fairness. Nonetheless, it is valuable to also delve into the existing literature concerning utility tokens.

B.2. Utility token

From a user’s perspective, a utility token serves as a ticket to efficiently match the demand and supply among participants. The value of a utility token is derived from users’ transactional demand within the DAO, incorporating positive network externalities (Cong, Li, and Wang, 2021a). In Lommers, Xu, and Xu (2022), another source of utility tokens value may come from future expected free cash flow in DAOs. As for platforms, utility tokens offer a method to finance the establishment of profitable projects. Platforms have the ability to manage the supply of utility tokens through the issuance or token-burning processes Cong, Li, and Wang (2022a).

A prominent advantage associated with the utility token is its capacity to mitigate potential commitment problems compared to centralized platforms. It alleviates the underinvestment dilemma, fosters the attraction of an initial user adoption to the platform, and facilitates the platform's adoption of new marketplaces by avoiding later coordination failures between transaction counterparties during the platform operation (Bakos and Halaburda, 2022; Li and Mann, 2018; Sockin and Xiong, 2023a).

Nevertheless, there remain potential concerns associated with utility tokens. For instance, as discussed in Gan, Tsoukalas, and Netessine (2021), the initial coin offering (ICO) of utility tokens can give rise to notable agency costs, underproduction issues, and a diminution in firm value, whereas such questions appear to be less severe when it comes to governance token issuance. Furthermore, as outlined in Sockin and Xiong (2023b), the utility token market is susceptible to instability due to the interplay of network effects influencing user participation and the non-neutrality of token prices, which can potentially lead to market disruptions. Although token retradeability can help mitigate this risk, concerns still persist.

Within our study, our primary focus is on governance tokens, and we operate under the assumption that users acquire governance tokens without any cost. Our findings reveal that even in the absence of utility tokens, governance tokens alone are capable of addressing the coordination challenge and have the capacity to draw a greater number of initial users to the platform. Furthermore, governance tokens can serve as strong incentives for platform founders to actively engage in the development of the DAO.

B.3. Voting in DAOs

A pivotal element within DAO governance pertains to the process of voting. Appel and Grennan (2023a) analyzes a comprehensive dataset comprising 10,639 proposals from 151 different DAOs. They focus on evaluating the extent of decentralization in decision-making within DAOs. Notably, their finding that there is a decline in the concentration of decision-making authority in DAOs corresponds with a substantial upsurge in the number of proposals.

Some researchers have dedicated their efforts to studying the optimal voting systems within DAOs. Tsoukalas and Falk (2020) have observed that the current form of linear stake-weighted voting (LV), characterized by τ-weighting, tends to discourage truthful voting behavior and diminish the platform's predictive accuracy. This voting system fails to achieve the first-best accuracy comparable to a centralized platform. Quadratic Voting (QV) has been proposed as a possible solution to LV. However, Benhaim, Hemenway Falk, and Tsoukalas (2023a) find that QV can exhibit both superior and inferior performance compared to LV. Both QV and LV can approach optimality asymptotically, especially as the number of voters tends towards infinity. In summary, enhancing voting efficiency within the platform may necessitate efforts to boost voters' participation.

Voting, when combined with various other mechanisms, can have a significant impact on the evolution of DAOs. In Benhaim, Hemenway Falk, and Tsoukalas (2023b) a committee-based consensus, coupled with approval voting, can yield substantial efficiency and resilience benefits to DAO, particularly when a sufficient number of voters actively participate. Furthermore, in Han, Lee, and Li (2023), a token-based voting system, combined with strategic token trading, can stimulate the growth of the platform by increasing the dispersion of token distribution.

In this paper, our emphasis lies in utilizing a straightforward voting system in conjunction with a buy-back regime. The assumption is that the voting system serves as a means for users to express their opinions and govern the utilization of their data.

III. Model Setup

A. *Platform in a data-driven production economy*

There are two periods and three dates $t \in \{0, 1, 2\}$. An economy contains a platform founder, a platform, and a continuum of potential users of the platform with a total measure of one. Potential users differ in the extent of privacy concerns associated with the extent of their personal data that is required to be shared with the platform. This heterogeneity is captured by $\theta \in [0, 1]$, where $\theta = 0$ ($\theta = 1$) denotes a user who suffers the highest (lowest) cost from sharing data with the platform among all potential users. The source of this heterogeneity may be attributed to factors such as users' age, with older individuals exhibiting greater reluctance to share data (Goldfarb and Tucker, 2012), or simply a utility primitive that comes from user's "taste" for privacy (Lin, 2022). At $t = 0$, each potential user is endowed with physical capital k. There is no depreciation on k for simplicity.

Production on the platform takes place in both periods, requiring users' personal data and capital as inputs. Per-period platform's production function is $Y_t = f(A_{t-1}, n_t, k) = A_{t-1}^{\eta} n_t k$, where Y_t with $t \in \{1, 2\}$ denotes the production in the first and second period. A_{t-1} denotes the requirement on the extent of data that users share with the platform at the beginning of each period. $n_t \in [0, 1]$ denotes the mass of users who joined the platform—user adoption. Following Jones and Tonetti (2020), we interpret η to be the importance of data as an input to production. We assume that platform's output increases with the extent of data shared, $f_A > 0$, user adoption, $f_n > 0$, and capital provided, $f_k > 0$. Platform's production function exhibits constant returns to scale in terms of user adoption $y_t = \frac{Y_t}{n_t} = A_{t-1}^{\eta} k$.

At date $t = 0$, the platform's initial extent of data required for the first-period production $A_0 \in (0, \overline{A}]$ is an exogenous platform characteristic and common knowledge. We interpret A_0 as the minimum requirement on the extent of data sharing that makes a given type of platform viable. A high A_0 indicates a platform that requires a large extent of users' data to be shared to produce goods, while a low A_0 indicates a platform that relies on little users' data to produce. Later, we will classify platforms into '*data-light*' and '*data-heavy*' types based on A_0. $\overline{A} > 1$ is the upper limit on the data allowed to be shared and is determined by laws/regulations that are exogenous to the platform. For example, if the government wants to protect users' welfare by limiting the extent of data sharing, it could set a low $\overline{A}$ to limit platforms' founders' ability to extract surplus from users. We analyze such policy intervention in Section V.E.

B. Decentralized governance of the platform

The platform's founder designs governance of the platform by setting up governance rules at date $t = 0$. Once the rules are set up, they become common knowledge and are automatically executed at date $t = 1$, that is, the platform is fully committed to its governance rules.

After observing platform's governance rules, the value of A_0, and the importance of data in production η at date $t = 0$, potential users decide whether to participate in the first-period platform production by contributing their capital k and sharing their data. At the end of the first period, users consume an equal share of the platform's output, and each user also receives one unit of a governance token as a reward for joining the platform. A user's utility depends on the consumption amount, the cost incurred due to data sharing, and a payoff from the possible sale of the governance token, if token buy-back is allowed. At each period t, only users with type $\theta < A_{t-1}$ who join the platform incur the cost. Potential users who do not join the platform cannot produce. They do not consume, and their utility is normalized to zero.

After the first-period consumption takes place and governance tokens are distributed at date $t =$

1, the platform automatically issues a proposal $0 \leq a \leq \overline{A} - A_0$, where a is the increase in the extent of data sharing required. The extent of data sharing for the second-period platform production is thus $A_1 = A_0 + a$ with range $[A_0, \overline{A}]$. Users with governance tokens vote 'yes' or 'no' on the proposal. For the proposal to pass, $\tau \in [0, 1]$ is the fraction of token holders who need to vote 'yes'. If $\tau = 1$, the voting requires everyone to agree. If $\tau = 0$, only the lowest-cost user $\theta = 1$ matters.[9] If not enough users agree with the proposal, it fails. After a proposal fails, the platform issues a new proposal with $a - \epsilon$, with $\epsilon > 0$ infinitely small. The platform repeats the procedure until a proposal passes. For simplicity, we assume that the platform directly proposes the highest a that just passes the voting rule threshold.

There are two possible governance rules regimes denoted by triplet (Γ, τ, p). (i) *Voting-only regime*, $(\Gamma = 1, \tau, 0)$: Users' only governance choice is to exercise their voting rights. If the proposal passes, A_1 becomes common knowledge and all potential users decide whether to join the platform in the second period. (ii) *Buy-back and voting regime*, $(\Gamma = 2, \tau, p)$: Platform provides an option for governance token holders to sell their tokens to the platform at price $P = p\,k$, where $p \geq 0$. Users who sell their governance tokens lose their voting rights and the right to participate in the second-period platform production. Users who decide not to sell their governance token vote to decide about the proposal and they have to divert part of the platform's output to finance the buy-back of tokens. Tokens are burned after buy-back. In the second period, the platform produces goods with the extent of data requirement A_1. At date $t = 2$, users share the platform's output, make payments for the buy-back, and then consume. Users with type $\theta < A_1$ who join the platform incur the cost. Users who sell their governance tokens consume $p\,k$.

C. Platform users' utility and founder's objective function

In each period t, if a user with type θ participates in the platform, he incurs cost C_t due to privacy concerns with the cost function:

$$C_t(A_{t-1}; \theta) = \mathbb{1}\{\theta < A_{t-1}\} \frac{A_{t-1} - \theta}{\overline{A} - A_{t-1}}. \tag{1}$$

[9]For simplicity, we assume that users have equal voting power. In reality, there is a complex distribution of voting power among users, and we thus interpret τ as the fraction of users needed to vote 'yes' instead of the number of 'yes' votes. For this reason, we allow $\tau < 0.5$ since it may be that a small number of users may have more than half voting power, see Appel and Grennan (2023a).

The user only incurs cost if his type $\theta < A_{t-1}$. The cost is approaching infinity when A_{t-1} approaches $\overline{A}$. A potential user's utility is defined as

$$U(c_1, c_2, A_1; \theta, A_0, \Gamma, \tau, p) = U_1 + \beta\, U_2, \tag{2}$$

where U_t denotes the utility in the first or second period and $\beta \in (0, 1]$ is a discount factor.

In the first period, the user's utility from joining the platform is $U_1 = c_1 - C_1(A_0; \theta)$, where $c_1 = y_1 = A_0^{\eta} k$ is the user's first-period share of the platform's output. In the second period, the user's utility from participating in the platform is $U_2 = c_2 - C_2(A_1; \theta)$, where c_2 is the user's second-period consumption. Consumption c_2 depends on whether the user sells his governance token or not. If the user does not sell, c_2 is the platform's output per user minus (per platform user) the amount spent on the buy-back of tokens. If the user sells the governance token, his second-period utility is $U_2 = p\, k$. If a potential user does not participate in the platform in either of the two periods, his utility is $U_t = 0$. Since we focus on the interaction of platform governance and production, we abstract away from any transaction cost users would have from interacting with the platform.

Platform founder's objective is to choose governance rules of the platform (Γ, τ, p) at $t = 0$ to maximize the platform's total output $Y = Y_1 + \beta\, Y_2$:

$$(\Gamma^*, \tau^*, p^*) = argmax_{\Gamma \in \{1,2\};\, \tau \in [0,1];\, p \in [0,\infty)}\, Y_1 + \beta\, Y_2. \tag{3}$$

A rationale for choosing this objective function is that, after the establishment of the platform, the founder's payoff is monotonically related to the volume of economic activities on the platform. For simplicity, we do not model this payoff. The reservation value of the platform's founder is zero. If the platform does not exist, all the agents in the economy have zero utility.

D. Centrally governed platform benchmark

To compare decentralized governance with platform governance that is organized according to investor-owned firm tradition, we introduce a model of a centrally governed platform. In this case, the platform's founder "owns" the platform, which means that, in each period, users' only decision is to choose to participate in the platform or not. The model timing and production technology are as in Section III.A. The platform owner's objective is to maximize the platform's total output $Y_1 + \beta Y_2$.

In the first period, given the initial extent of data requirement A_0 and the importance of data in production η, potential users decide whether to contribute their capital k to the platform and participate in the production process by sharing their data. The platform production function is $f(A_0, n_1, k) = A_0^\eta n_1 k$, where n_1 is the first-period platform adoption. A user consumes $c_1 = A_0^\eta k$ and incurs cost $C_1(A_0; \theta)$ from participating in the platform. Different from the decentralized governance case of Section III.B, at date $t = 1$, the platform owner unilaterally decides the incremental data required by the platform, $a \in [0, \overline{A} - A_0]$, and thereby the extent of data sharing required in the second-period platform production, $A_1 = A_0 + a$.

In the second period, the platform production function is $f(A_1, n_2, k) = A_1^\eta n_2 k$, where n_2 is the second-period platform adoption. Users observe A_1 and decide whether to participate in the platform and incur cost $C_2(A_1; \theta)$. Each user could receive $c_2 = A_1^\eta k$ as his second-period consumption, however, users do not have control rights over the platform and the right to distribute the second-period output resides with the platform owner. Although, at $t = 0$, the platform owner may promise to users that they will receive $c_2 = A_1^\eta k$ in the second period, platform owner cannot commit to provide this consumption. The platform owner may violate her promise and carry out a transfer policy in the second period. For example, she may reduce consumption for users with high θ and transfer this consumption to subsidize users with low θ. This transfer policy $c^{\text{transf}}(\theta)$ attracts users with low θ into the platform which increases Y_2.

In the centrally governed platform benchmark, the platform owner's problem is:

$$max_{A_1 \in [A_0, \overline{A}];\, c^{\text{transf}}(\theta)} \; Y_1 + \beta Y_2. \tag{4}$$

E. Social planner platform governance benchmark

In addition to the platform organized by its founder as decentralized or centrally governed, we examine the social planner's approach to platform governance. We assume that the social planner has complete control over the platform, including decisions about user participation in both periods. The model timing and production technology are as in Section III.A. The key distinction from the platform founder's approach is that the social planner's objective is to maximize social welfare, which, in our context, equates to user surplus.

Given the platform's initial extent of data required A_0 and the importance of data in production η, the social planner decides what type of users should contribute their capital k and participate in the platform production in the first period. In the first period, the platform production function is $f(A_0, n_1, k) = A_0^\eta n_1 k$, where n_1 is the first-period platform adoption decided by the social planner. A user consumes $c_1 = A_0^\eta k$ and incurs cost $C_1(A_0; \theta)$ from participating in the platform. Different from the decentralized governance case of Section III.B, at date $t = 1$, the social planner decides the incremental data required by the platform, $a \in [0, \overline{A} - A_0]$, and thereby the extent of data sharing required in the second-period platform production, $A_1 = A_0 + a$. In the second period, the platform production function is $f(A_1, n_2, k) = A_1^\eta n_2 k$, where n_2 is the second-period platform adoption, again, decided by the social planner. A user consumes $c_2 = A_1^\eta k$ and incurs cost $C_2(A_1; \theta)$ from participating in the platform.

We define per-period users' surplus S_t as:

$$S_t = \int_{\theta_t}^{1} U_t(\theta) d\theta, \tag{5}$$

where θ_t is the lowest user type that joins the platform in period t and a user's utility from participating in the platform is $U_t = c_t - C_t(A_{t-1}; \theta)$. Compared to the total output objective, users' surplus takes privacy concerns of users directly into consideration. In our definition of users' surplus as a proxy for social welfare, we ignore the utility of the platform's founder. This is because, in the model, the total measure of all potential users of the platform is one while the platform founder alone is zero. Therefore, the users' surplus is equal to the average surplus in the economy.[10]

The social planner's objective is to maximize the total users' surplus $S = S_1 + \beta\, S_2$:

$$max_{\theta_1 \in [0,1];\ \theta_2 \in [0,1];\ a \in [0, \overline{A} - A_0]}\ S_1 + \beta\, S_2 = max_{\theta_1 \in [0,1]}\ S_1 + max_{\theta_2 \in [0,1];\ a \in [0, \overline{A} - A_0]}\ \beta\, S_2, \tag{6}$$

[10]We assume that the platform's founder aims to maximize the total output of the platform, as her payoff increases with the volume of transactions on the platform. This implies that we study a self-interested founder who is indifferent to the welfare of users. Although we do not explicitly model the founder's payoff, incorporating it would not alter the overall findings of this paper. Purposefully, the social planner has a markedly different objective. Furthermore, we note that after the production process on the platform is completed, the social planner can determine how to distribute the total surplus, however, examining such a redistribution policy is beyond the scope of this paper.

where θ_1 and θ_2 are defined as the lowest user type that joins the platform in the first or second period, respectively. Since A_0 is exogenous, the decisions made by the social planner are separable by period.

IV. Optimal Decentralized Governance

In this section, we present solutions to different governance regimes in Section III.B. We focus on $\eta = 1$ case, where data is equally important as the real capital owned by the users. The discussion of $\eta > 1$ case is provided in the online appendix. For simplicity, we normalize k to 1. In this case, the output is in the unit of initial capital, and the buy-back price corresponds to the real price. Therefore, the new production function is given by $f(A_{t-1}, n_t, k) = A_{t-1}n_t$, and the objective of the decentralized platform founder is to maximize the total output, which is defined as $Y_1 + \beta Y_2 = A_0 n_1 + \beta A_1 n_2$.

A. *Properties of the cost and utility functions*

To begin with the model solution, it is useful to examine some properties of the cost and utility functions.

Lemma 1: Properties of cost function. The cost function $C_t(A_{t-1};\theta) = \mathbb{1}\{\theta < A_{t-1}\}\frac{A_{t-1}-\theta}{\bar{A}-A_{t-1}}$ is (1) a weakly increasing function of A_{t-1} for $\forall\theta$; (2) a weakly decreasing function of θ for $\forall A_{t-1}$. (3) When $\theta < A_{t-1}$, the derivatives of cost function $C_t(A_{t-1};\theta)$ with respect to A_{t-1} is a decreasing function of θ.

Lemma 1 implies that, for any A_{t-1}, users' costs are ranked in the order of their type. User with $\theta = 0$ has the highest costs, while user with $\theta = 1$ has the lowest costs. Consequently, users' utility functions from joining the platform and excluding buy-back payments exhibit similar properties.

Lemma 2: Properties of users' utility function from joining the platform. (1) The utility function $U_t(A_{t-1};\theta) = A_{t-1} - \mathbb{1}\{\theta < A_{t-1}\}\frac{A_{t-1}-\theta}{\bar{A}-A_{t-1}}$ is a weakly increasing function of θ for $\forall A_{t-1}$. (2) If $U_t(A_L;\theta) \le 0$, then $U_t(A_H;\theta) \le 0$ for any $A_H \ge A_L$.

The first property in Lemma 2 is a direct result of Lemma 1. The second property illustrates that if users are unwilling to join the platform in the first period ($U_1(A_0;\theta) \leq 0$), they will not join the platform in the second period since $A_1 \geq A_0$, which implies that the second-period utility is also negative.

Figure 1 displays a graph of the user's cost and utility as functions of the requirement on the extent of data sharing A_{t-1} when $\theta = 0.4$ or $\theta = 0.6$. For Panel A, the cost is zero when $A_{t-1} \leq \theta$ and approaches infinity when A_{t-1} is close to $\overline{A}$. In Panel B, user's utility first increases and then decreases to negative infinity, since the privacy cost dominates the data-sharing benefit when $A_{t-1} \to \overline{A}$. The proof is available in the appendix.

B. Solution of the voting-only decentralized governance regime

According to the model's setup, the minimum (initial) requirement on the extent of data sharing A_0 and the voting regime are common knowledge for users in the economy. The decentralized platform owner chooses the voting threshold $\tau \in [0,1]$ to maximize the total output $Y_1 + \beta Y_2$ in both the first and second periods.

We define θ_1 and θ_2 to be the lowest type users who participate in the platform production during the first and second periods, respectively. According to Lemma 2, in both periods, all the users with $\theta \geq \theta_t$ will also join the platform. User adoption in both periods can be obtained by $n_t = 1 - \theta_t$.

We employ a backward induction approach to solve the problem. First, in the second period, given the voting result in a and the corresponding data sharing requirement $A_1 = A_0 + a$, all the agents with type $\theta \in [0,1]$ in the economy are faced with the decision of whether to participate in the platform or abstain. Participating users evenly distribute the second-period output, which suggests $c_2 = \frac{Y_2}{n_2} = A_1$. User utility is expressed as $U_2 = A_1 - \mathbb{1}\{\theta < A_1\}\frac{A_1-\theta}{\overline{A}-A_1}$. Users with a utility of $U_2 \geq 0$ opt for participation. A special user θ with utility $U_2 = 0$ is determined by the equation below:

$$A_1 - \mathbb{1}\{\theta < A_1\}\frac{A_1 - \theta}{\overline{A} - A_1} = 0$$

Solving the equation above, we find that the root is $A_1 - A_1(\overline{A} - A_1)$. It is possible that this root is negative. If $\overline{A} - A_1 \leq 1$, the lowest type of user participating in the production process is expressed as $\theta_2 = A_1 - A_1(\overline{A} - A_1)$. However, if $\overline{A} - A_1 > 1$, all potential users are willing to join the platform, and $\theta_2 = 0$. Therefore, the expression for θ_2 can be summarized as follows:

$$\theta_2 = max\{0, A_1 - A_1(\overline{A} - A_1)\} \tag{7}$$

For the voting process, at $t = 1$, a group of users with type $\theta \in [\theta_1, 1]$ engage in the first-period production and each acquires a governance token. Possessing these tokens confers voting rights upon them, which they will exercise in the decision-making process regarding the proposed value of a. Each user, characterized by their unique type θ, endeavors to optimize their second-period utility. If a user with type θ is able to independently determine A_1, his problem becomes:

$$max_{A_1 \in [A_0, \overline{A}]} A_1 - \mathbb{1}\{\theta < A_1\} \frac{A_1 - \theta}{\overline{A} - A_1}$$

We denote A^*_θ as the optimal solution for the problem above and $U^*_{2,\theta} = U_2(A^*_\theta)$ represents the maximized second-period utility for a user with type θ. Solving the equation above, the optimal second-period data sharing for the user with type θ is $A^*_\theta = \overline{A} - \sqrt{\overline{A} - \theta}$. Then, we have the following lemma:

Lemma 3: Properties of user's optimal second-period utility. (1)$\overline{A} > A^*_\theta \geq \theta$. (2) A^*_θ is increasing in θ . (3) $U^*_{2,\theta}$ is weakly increasing in θ.

According to Lemma 3 and the platform voting setting, we can obtain the voting strategy for the user with type θ.

Proposition 1: The voting strategy of the user with type θ in the voting-only regime The optimal voting strategy for the user with type θ is to vote 'yes' for $\forall a \leq A^*_\theta - A_0$ and 'no' for $\forall a > A^*_\theta - A_0$.

A key intuition underlying Proposition 1 is that, if a user with type θ votes 'no' for the current proposal satisfying $a \leq A^*_\theta - A_0$, his second-period utility may deteriorate. If the current proposal fails, a subsequent proposal with a lower value of a will be introduced and the second proposal will deviate more from the user's optimal. Conversely, if the user votes 'no' for a proposal with $a > A^*_\theta - A_0$, he can anticipate a higher second-period utility if the current proposal fails to pass.

Proposition 1 also suggests that if a user with type θ agrees with the proposal, all users with higher θ values will also concur with the proposal. Consequently, given the voting rule τ, the pivotal user in this voting process is the type $\theta_{pivotal} = 1 - n_1\tau$. For the pivotal user, his optimal second-period data-sharing requirement is denoted as $A^*_{\theta_{\text{pivotal}}}$. If $A^*_{\theta_{\text{pivotal}}} \geq A_0$, to streamline the process, the platform will propose $a^* = A^*_{\theta_{\text{pivotal}}} - A_0$ to ensure that the voting result just reaches the threshold. However, in cases where the initial data requirement is exceedingly high $A^*_{\theta_{pivotal}} < A_0$, the platform will refrain from proposing any value, leading to $a = 0$. In summary, the optimal data-sharing in the second period is:

$$A^*_1 = max\{A_0, A^*_{\theta_{pivotal}}\} = max\{A_0, \overline{A} - \sqrt{\overline{A} - \theta_{pivotal}}\} \tag{8}$$

At $t = 0$, all users within the economy make decisions regarding their participation in the platform's first-period production. The crucial type to consider is θ_1 which is the lowest user type willing to join in the platform. Users who choose to participate will evenly distribute the platform's output, $c_1 = \frac{Y_1}{n_1} = A_0$. A user will opt to join the platform in the first period if:

$$U(c_1, c_2, A_0, A_1) = U_1 + \beta U_2 \geq 0$$

By utilizing the second property outlined in Lemma 2, we can obtain the characteristics of first-period user adoption within the voting-only regime. This implies that the voting-only regime does not have the ability to attract additional users to the platform initially. Under this setting, only users experiencing non-negative utility will choose to participate in the platform in the first period. Additionally, it is worth noting that second-period user adoption never exceeds first-period user adoption $n_2 \leq n_1$. The lowest θ with utility $U_1 = 0$ is determined by the equation below:

$$A_0 - \mathbb{1}\{\theta < A_0\}\frac{A_0 - \theta}{\overline{A} - A_0} = 0$$

Solving the equation above, we obtain the root $A_0 - A_0(\overline{A} - A_0)$. The root is below zero if $\overline{A} - A_0 > 1$. On the contrary, the root is larger than zero if $\overline{A} - A_0 \leq 1$. Therefore, the expression of θ_1 is:

$$\theta_1 = max\{0, A_0 - A_0(\overline{A} - A_0)\} \tag{9}$$

The user participation dynamics throughout the entire voting process can be simplified as follows: In the first period, users with types $\theta \in [\theta_1, 1]$ participate in the platform for production. Then, at $t = 1$, users belonging to the type $\theta \in [\theta_1, 1]$ cast their votes regarding the proposal a. In the second period, users within the type $[\theta_1, \theta_2]$ exit the platform, while the rest $\theta \in [\theta_2, 1]$ continue to remain within the platform for production.

We then shift our focus to the platform founder's problem. Given $p = 0$, the optimal voting threshold τ^* is the one maximizes the total social output $Y_1 + \beta Y_2$ in both periods:

$$\tau^* = argmax_{\tau \in [0,1]}\ A_0 n_1 + \beta\, A_1 n_2 \tag{10}$$

A_0 is exogenous and describes the minimum (initial) requirement on the extent of data sharing that makes a platform viable. At the same time, $n_1 = 1 - \theta_1 = 1 - max\{0, A_0 - A_0(\overline{A} - A_0)\}$ is a function of A_0 and is not affected by τ. β is a constant term. Hence, the problem described above is equivalent to the maximization of second-period output:

$$\tau^* = argmax_{\tau \in (0,1]}\ A_1 n_2$$

In the case of the voting-only regime, the sole parameter under the control of the platform founder is the voting threshold τ. Altering the voting threshold τ directly impacts second-period data sharing A_1^* and indirectly influences second-period user adoption n_2.

Specifically, an increase in τ results in a decrease in the pivotal voter's type $\theta_{\text{pivotal}} = 1 - n_1\tau$, subsequently reducing A_1^*. The rise of A_1 has two consequences: it raises the average output which is the benefit of sharing-data $c_2 = \frac{Y_2}{n_2} = A_1$, and users' privacy cost $\mathbb{1}\{\theta < A_1\}\frac{A_1 - \theta}{\overline{A} - A_1}$. Whether user adoption will increase depends on which force dominates. Figure 2 gives an example of how changes in τ influence both A_1 and n_2. As τ increases from zero, second-period data sharing A_1

decreases till it reaches the minimum. Simultaneously, the data-sharing benefits dominate and the platform attracts more users in the second period.

B.1. Comparison of 'data-heavy' platform vs. 'data-light' platform

A_0 is the initial extent of data required by the platform to produce. It describes the characteristics of the platform, a higher A_0 indicates that the goods produced by the platform are data-intensive. Whether the platform is '*data-light*' or '*data-heavy*' depends only on A_0 and $\overline{A}$.

If $A_0 \leq A_{max}$, we define the platform to be a '*data-light*' platform. If $A_0 > A_{max}$, the platform is a '*data-heavy*' platform, where A_{max} is expressed as:

$$A_{max} = \begin{cases} \overline{A} - \sqrt{\overline{A} - 1} & \text{if } 2 < \overline{A} \\ \overline{A} - 1 & \text{if } \overline{A}_{root} < \overline{A} \leq 2 \\ \frac{\overline{A} - 1 + \sqrt{(\overline{A}-1)^2 + 3}}{3} & \text{if } 1 \leq \overline{A} \leq \overline{A}_{root} \end{cases}$$

Denote $\frac{\overline{A}-1+\sqrt{(\overline{A}-1)^2+3}}{3} = A_{max1}$ for simplicity, $\overline{A}_{root}$ is the root of the following equation:

$$A_{max1}[1 - (A_{max1} - A_{max1}(\overline{A} - A_{max1}))] - (\overline{A} - 1) = 0$$

we have the following lemma:

Lemma 4: Properties of '*data-light*' and '*data-heavy*' platform. Given $\overline{A}$, in the '*data-light*' platform, the second-period data-sharing requirement is $A_1^* = A_{max}$. User adoption is determined by $n_2 = 1 - \theta_2 = 1 - (A_{max} - A_{max}(\overline{A} - A_{max}))$. In the '*data-heavy*' platform, the second-period data-sharing requirement is $A_1^* = A_0$. User adoption is determined by $n_2 = 1 - \theta_2 = 1 - (A_0 - A_0(\overline{A} - A_0))$.

Lemma 4 implies that, for second-period production, there is an optimal $A_1^* = A_{max}$ that maximizes second-period output. A_1^* is achievable when A_0 is small (in '*data-light*' platforms). When $A_0 > A_{max}$ in '*data-heavy*' platforms, since we require $A_1^* \geq A_0$, A_{max} is not achievable.

Figure 3 illustrates how the voting threshold τ, total output $Y_1 + \beta Y_2$, total user surplus $S_1 + \beta S_2$, second-period data sharing A_1 and user adoption n_1 and n_2 changes with the first-period data sharing requirement A_0 in the optimal.

According to the graph, first-period user adoption n_1 and second-period user adoption n_2 is non-increasing in A_0. This indicates that although A_0 increases the output, the negative effects on the user's cost dominate and decrease user adoption. A_1 is always non-decreasing in A_0. The optimal voting threshold is not monotonic with A_0. When $\overline{A} = 1.05$, which means that the privacy costs are relatively high for all users, we see that in a '*data-light*' firm ($A_0 < 0.594$), the optimal voting threshold is set to maintain a steady A_1 and n_2. In a '*data-heavy*' firm($A_0 \geq 0.594$), the optimal voting threshold is set to push the second-period data sharing as high as possible. Besides, the total output and total user surplus first increase and then decrease with A_0.

C. *Solution of buy-back and voting decentralized governance regime*

In this setting, the minimum (initial) requirement on the extent of data sharing A_0 and the buy-back and voting regime are common knowledge for users in the economy. The decentralized platform owner chooses the voting threshold τ and buy-back price ratio p to maximize the total output $Y_1 + \beta Y_2$ in both periods.

The same methodology is applied to solve the problem of the buy-back and voting regime. Similar to the previous Section IV.B, we define θ_1 to be the lowest type that engages with the platform during the first period. Additionally, we introduce $\theta_{sell} \geq \theta_1$ to represent the type that is indifferent between selling the governance token and retaining their platform position. Furthermore, θ_2 is the lowest type participating in the platform's second-period production. The first-period and second-period user adoption is $n_1 = 1 - \theta_1$ and $n_2 = 1 - \theta_1$, respectively. Finally, we denote $n_{sell} = 1 - \theta_{sell}$ as the mass of users who have voting power and actively vote for the proposal.

First, in the second period, given the voting outcome a and the corresponding data-sharing $A_1 = A_0 + a$, all the users with type $\theta \in [0, \theta_1]$ and $\theta \in [\theta_{sell}, 1]$ in the economy make a choice regarding their participation in the platform. Users with type $[\theta_1, \theta_{sell}]$ have sold their governance tokens and are not permitted to partake in the second-period production. Participating users equally divide the second-period output, receiving $\frac{Y_2}{n_2} = A_1$.

Following the distribution of the output, users with type $\theta \in [\theta_{sell}, 1]$ are obligated to make payments to the sellers of governance tokens. The amount of payment equals $\frac{\theta_{sell} - \theta_1}{1 - \theta_{sell}} p$ and and it is identical across all payers. On the contrary, users with type $[\theta_1, \theta_{sell}]$ receive p. Subsequently, everyone proceeds to consume, and their respective utility is realized. Nevertheless, it is important to note that the buyback payoff for the seller and the payment for the payee do not influence the second-period participation decision, as the buyback decision has already been finalized.

An important implication here is that after the buyback process, users holding governance tokens are bound to participate in the platform during the second period. Refraining from participation in the second period would result in a utility of zero, which is less favorable than the alternative of selling their governance tokens during the buy-back process and receiving a payment of p. Furthermore, according to lemma2 (2), users with type $\theta \in [0, \theta_1]$ will not engage in second-period production, as their participation would result in negative utility across both time periods. Consequently, the problem of second-period participation becomes exceedingly straightforward, which is :

$$\theta_2 = \theta_{sell} \tag{11}$$

Subsequently, for the voting process, a group of users with type $\theta \in [\theta_{sell}, 1]$ will vote for the proposed a. Taken θ_1 and θ_{sell} as given, for each user θ, his optimal A^*_θ is obtained by:

$$A^*_\theta = argmax_{A_1 \in [A_0, \overline{A}]} \, A_1 - \mathbb{1}\{\theta < A_1\} \frac{A_1 - \theta}{\overline{A} - A_1}$$

According to lemma 3 and proposition 1, the type θ user's voting strategy is that he will agree if the proposed $a \leq A^*_\theta - A_0$ and not agree if the proposed $a > A^*_\theta - A_0$.

Similar to Section IV.B, with the voting rule τ, the pivotal user is denoted as $\theta_{\text{pivotal}} = 1 - n_{\text{sell}}\tau$. Consequently, the optimal proposal $A^*_{\theta_{\text{pivotal}}}$ is the one that maximizes his utility function. Nevertheless, it remains necessary to compare the pivotal voter's optimal solution with the initial data-sharing requirement A_0. In summary, the requirement of data sharing in the second period can be expressed as:

$$A_1 = max\{A_0, A^*_{\theta_{pivotal}}\} = max\{A_0, \overline{A} - \sqrt{\overline{A} - \theta_{pivotal}}\} \tag{12}$$

In contrast to Section IV.B, the buy-back process permits users with types $\theta \in [\theta_1, 1]$ to decide whether to sell their governance tokens at price p. To achieve market clearance, it is necessary that the price p falls within a specific range. If p is excessively high or excessively low, it may result in a situation where no one in the market is willing to sell their governance token, or there is simply no demand for the governance token. In the context of the buy-back process, we present the following proposition:

Proposition 2: Buy-back market. For any τ and A_0 there exists some buy-back price ratio $p > 0$ clear the buy-back market. As long as the market is clear, there exists a special type θ_{sell} such that: 1) User with type θ_{sell} is indifferent between sell and not sell and his second-period utility is p. 2) For any $\theta_{sell} \geq \theta \geq \theta_1$, users' second-period utility are equal to p.

For any users with type $1 \geq \theta \geq \theta_{sell}$, they choose to stay in the platform to produce and equally share the output. They also need to pay $p(\theta_{sell} - \theta_1)$ in total to those who sell the governance token. Therefore, buy-back is indeed a transfer from low-cost users to high-cost users.

$\theta_{sell} \in [\theta_1, 1]$ can be obtained by the equation below:

$$p = A_1^* - \mathbb{1}\{\theta_{sell} < A_1^*\}\frac{A_1^* - \theta_{sell}}{\bar{A} - A_1^*} - \frac{n_1 - n_{sell}}{n_{sell}}p$$

Rewrite the equation, the price of buy-back can be written as:

$$p = \frac{n_{sell}}{n_1}(A_1^* - \mathbb{1}\{\theta_{sell} < A_1^*\}\frac{A_1^* - \theta_{sell}}{\bar{A} - A_1^*}) = \frac{n_{sell}}{n_1}U_2(A_1^*; \theta_{sell}) \tag{13}$$

This implies that the buyback price is equivalent to the second-period utility of a user with type θ_{sell} before the buyback payment, and then multiplied by the remaining users' ratio.

At $t = 0$, users make decisions regarding their participation in the platform during the first period. The crucial type to consider is $\theta_1 = 1 - n_1$. Under the buy-back and voting regime, first-period user adoption differs from what was discussed in Section IV.B. For any price ratio p that clears the buyback market, there will be users who join the platform experiencing a negative first-period utility. This suggests that, under the buy-back regime, the platform can attract additional users with a guaranteed payoff p since they receive a governance token, which protects their rights during the buy-back process.

Provided that the buyback price ratio clears the market, there exists an individual who is indifferent between engaging in first-period production then subsequently selling the governance token, and abstaining from participation in the platform in both periods. This situation can be characterized by the following equation:

$$0 = A_0 - \mathbb{1}\{\theta < A_0\}\frac{A_0 - \theta}{\overline{A} - A_0} + \beta\, p$$

Solving the equation above, we obtain the root $A_0 - (A_0 + \beta\, p)(\overline{A} - A_0)$. Therefore, we have:

$$\theta_1 = max\{0, A_0 - (A_0 + \beta\, p)(\overline{A} - A_0)\} \tag{14}$$

The user participation dynamics during the entire voting and buyback process can be summarized as follows: In the first period, users with types $\theta \in [\theta_1, 1]$ engage in platform-based production. Then, at $t = 1$, users within the type range $\theta \in [\theta_1, \theta_{\text{sell}}]$ sell their governance tokens to users with types $\theta \in [\theta_{\text{sell}}, 1]$. Following this transaction, users with types $\theta \in [\theta_{\text{sell}}, 1]$ determine the incremental data sharing requirement a. In the second period, users with types $\theta \in [\theta_{\text{sell}}, 1]$ continue to participate in platform-based production.

Back to the platform founder's problem, the optimal voting rule threshold and buy-back price τ^*, p^* is the pair of values that maximizes the total output $Y_1 + \beta\, Y_2$ in both periods:

$$(\tau^*, p^*) = argmax\ A_0 n_1 + \beta\, A_1 n_2 \tag{15}$$

Similar to Section IV.B, the voting threshold τ does not impact first-period total output. However, it directly affects A_1^* and indirectly influences n_2. The buy-back price ratio p will have impacts on both periods. Firstly, first-period user adoption n_1 is increasing in p. Secondly, the increase in p has a dual effect on θ_{sell} and consequently indirectly influences A_1^* and n_2. On one hand, a higher buy-back price encourages the participation of low-cost users during the first period. This leads to an increase in n_1 and the components of users possessing governance tokens. On the other hand, the rise in p fosters a greater incentive to sell the governance token. Whether θ_{sell} increases or decreases depends on which of these forces dominates.

Figure 4 gives an example of how user's participation in voting n_{sell}, second-period user adoption n_2, and second-period data sharing A_1 change with the buy-back price ratio p. In the graph, fixed τ

and A_0, we could see the n_{sell} first increase then decrease when p increases, which indicates that the attracting users effect first dominates and then the willingness to sell effects dominates. Similarly, the effects of p on A_1 and n_2 are not monotone.

The comparison between the voting-only regime and buy-back and the voting regime is discussed in Section IV.D.

C.1. Comparison of 'data-heavy' platform vs. 'data-light' platform

For the buying back and voting case, the platform founder is able to determine the voting rule threshold τ and the buying back price ratio p.

The optimal solution in buy-back and the voting regime is presented in Figure 5. First, the buy-back market collapses in the '*data-light*' platforms. The optimal voting threshold is zero to achieve $A_{max} = 1$. Second, in '*data-heavy*' platforms, a buy-back market exists when A_0 is not close to $\overline{A}$, since in this case the privacy cost is high and everyone wants to sell the governance token. When buy-back market exists, the optimal buy-back price and voting threshold present an inverse "U-shape". p and τ first increase with A_0 and then decrease with A_0. Similar to figure 3, A_1 is always non-decreasing in and A_0. n_1, n_{sell}, and n_2 is non-increasing in A_0. Total output reaches its maximum in '*data-heavy*' platform while total surplus reaches its maximum in '*data-light*' platform.

D. Comparison of the two decentralized governance regimes

In this section, we compare the difference between the voting-only and buy-back regimes. First, in '*data-light*' platforms, the buy-back market does not exist. There is no difference between these two regimes.

In '*data-heavy*' platforms, if a buy-back market exists, first-period user adoption n_1 is always higher in the buy-back regime. Since A_0 is given, therefore, buy-back regime will always generate higher output in the first period. On the other hand, although a voting-only regime cannot attract more users to the platform, its optimal solution is the same as the social optimum solution in the first period. Therefore, the voting-only regime will always generate higher social welfare in the first period.

Figure 6 presents the difference on $n_1, n_2, A_1, a, Y_1 + \beta\, Y_2$, and $S_1 + \beta\, S_2$ between the voting-only regime and buy-back regime ($\overline{A}$). According to the second and third panels in figure 6, second-period user adoption n_2 is always higher in the voting-only regime. However, voting-only and buy-back and voting are able to achieve the same second-period data sharing A_1 in the simulation.

Also, for the voting-only case, in the simulation, its total output is always smaller than or equal to the buy-back case while its user surplus is always larger than or equal to the buy-back case. We have proposition 3 below as the key results.

Proposition 3: Choice of decentralized platform owner. In '*data-light*' platforms, the buy-back and voting regime is indifferent to the voting-only regime. In '*data-heavy*' platforms, the buy-back and voting regime weakly dominates the voting-only regime for the decentralized platform owner.

In all, since the decentralized platform owner only cares about total output, the best strategy is to implement a buy-back and voting regime. However, when the platform is '*data-light*', the platform owner is indifferent between these two regimes.

V. Comparison of Optimal Decentralized Governance to Centrally Governed and Social Planner Benchmarks

To illustrate the efficiency and fairness of the decentralized platform, we begin by presenting the results of two benchmarks: the centralized platform and the social planner's model. Following this, we will compare the total output and explain the incentives for the platform founder to establish a decentralized platform. We will also compare user surplus to demonstrate the fairness of the decentralized platform in terms of social welfare. Finally, we will discuss the policy where the government put restrictions on data sharing and the policy's potential impact on platform productivity.

A. *Solution of the centrally governed platform benchmark*

The platform's owner has full ownership control. Governance occurs at $t = 1$ when the owner determines the data sharing requirement A_1 for the second period. Users lack control rights over the centralized platform.

A commitment issue arises wherein the centralized owner can implement a transfer policy $c^{\text{transf}}(\theta)$ in the second period. Even if the owner promises users high output distribution in the second period, she can still renege on it. Upon learning of this possibility, users become cautious when deciding to join the platform. Consequently, a critical characteristic of first-period user adoption on the centralized platform is that users will only engage in first-period platform production if they anticipate a non-negative utility.

In the centralized firm, we denote θ_1 as the minimum type that opts to join the platform in the first period. The minimum (initial) requirement on the extent of data sharing A_0, is exogenously determined. Furthermore, the centralized platform owner cannot take any action in the first period as governance is only scheduled for $t = 1$. Therefore, same to the discussion in Section V.A, we have:

$$\theta_1 = max\{0, A_0 - A_0(\overline{A} - A_0)\} \tag{16}$$

which is equal to the voting-only regime.

In the second period, we define θ_2 to be the lowest type that joins the platform. The platform owner's behavior follows the proposition below:

Proposition 4: Second-period centralized owner's strategy. The platform owner will always implement a transfer policy $c^{\text{transf}}(\theta) = min\{-\int_0^{\underline{\theta}} U_2(\theta)d\theta, \int_{\underline{\theta}}^1 U_2(\theta)d\theta\}$ to redistribute output from low-cost users to high-cost users. $\underline{\theta}$ is the lowest type that join the platform with given A_1 when there is no interfere. As long as the centralized platform owner ensures that all users on the platform receive a non-negative utility, the transfer policy will increase the overall total output.

Accordingly, we have the following lemma for second-period user adoption.

Lemma 5: Second-period centralized platform user adoption. $\theta_2 = 1 - n_2$ equals to $\theta_2 = max\{0, \theta_{root}\}$. θ_{root} is the type that drives user surplus to zero and can be expressed as:

$$\theta_{root} = \frac{-E + \sqrt{E^2 - 4DF}}{2D}$$

where $D = -1$, $E = 2A_1 - 2A_1(\overline{A} - A_1)$, and $F = 2A_1(\overline{A} - A_1) + \mathbb{1}\{A_1 > 1\}(A_1 - 1)^2 - A_1^2$.

θ_{root} is determined by $S_2 = \int_{\theta_{root}}^{1} U_2(\theta)d\theta = 0$ so that the second period user surplus equals zero. Also, $U_2 = A_1 - \mathbb{1}\{A_1 > \theta\}\frac{A_1-\theta}{\overline{A}-A_1}$.

According to Lemma 5, second-period user adoption n_2 is a function of A_1. Platform owner maximizes the second-period output by choosing the optimal A_1:

$$max_{A_1 \in [A_0, \overline{A}]} \; A_1 n_2 \tag{17}$$

Figure 7 illustrates the centralized platform founder's problem of choosing A_1. For the owner, when A_1 increases, n_2 is non-increasing. Total output and user surplus first increase and then decrease. She will choose A_1 that maximizes the total output. However, the optimal A_1^* for a centralized platform owner does not maximize user surplus.

The founder or owner of the platform will seek to maximize overall output, whether the platform is centralized or decentralized. Our assumption is based on the fact that their payout rises as more transactions are made on the platform. The platform founder or owner does not care about the user's welfare. We do not model the platform's founder or owner's payoff function, but including their profit will not change the overall findings in this paper.

B. Solution of the social planner platform governance benchmark

In this part, we present the solution to the social planner's problem. Similar to the voting-only regime and the centralized platform, the social planner cannot affect the first-period data requirement A_0. The social planner tries to maximize user surplus in each period.

The following lemma provides some properties for the social planner's problem:

Lemma 6: Properties of social planner's optimal user adoption. For any given A_{t-1}, social planner chooses $\theta_t = max\{0, A_{t-1} - A_{t-1}(\overline{A} - A_{t-1})\}$.

Lemma 6 has several implications. First, the optimal solution for social planner is to prevent all negative utility users from participating in the platform. The solution can be realized without any interference. Second, given the same A_0, in the first period, the voting-only regime, the centralized platform, and the social planner model generate the same user adoption, output, and user surplus. Therefore, the only thing for the social planner to do is to choose an optimal A_1 to maximize second-period user surplus. The social planner's problem can be simplified as:

$$max_{A_1 \in [A_0, \overline{A}]} S_2$$

The expression of S_2 is $A_1(1-\theta) + \frac{1}{2(\overline{A}-A_1)}[\mathbb{1}\{A_1 > 1\}(A_1-1)^2 - \mathbb{1}\{A_1 > \theta\}(A_1-\theta)^2]$ with $\theta = max\{0, A_1 - A_1(\overline{A} - A_1)\}$.

Figure 8 illustrates the social planner's problem of choosing A_1. When A_1 increases, n_2 is non-increasing. Total output and user surplus first increase and then decrease. She will choose A_1 that maximizes the user surplus. However, one can clearly see that the optimal A_1^* for the social planner does not maximize output in the second period.

C. Total output ranking

In this section, we compared the total output of the decentralized platform and the benchmark. The following proposition illustrates the key results:

Proposition 5 Efficiency: Comparison on total output. Given the same A_0, the total output ($Y = Y_1 + \beta Y_2$) of the voting-only regime (VOTING), the buy-back and voting regime (BUY-BACK), the centralized platform (CENTRAL), and the social planner (SOCIAL) have the following rank:

$$Y_{SOCIAL} \leq Y_{VOTING} \leq Y_{CENTRAL} \tag{18}$$

$$Y_{SOCIAL} \leq Y_{VOTING} \leq Y_{BUY-BACK} \tag{19}$$

$$Y_{BUY-BACK} <> Y_{CENTRAL} \tag{20}$$

Figure 9 presents the total output ranking ($\overline{A} = 1$, $\beta = 1$). In this case, '*data-light*' and '*data-heavy*' platforms are distinguished by a threshold $A_0 = 1$. Panel A is the total outputs of benchmarks and voting-only as a function of A_0. When A_0 is extremely large, all three cases will have zero total output, which means no users are willing to participate in production. For a lower A_0 with non-zero user adoption, the centralized platform's output dominates voting-only and social planer all the time. The voting-only regime has a higher total output in '*data-light*' platforms compared to the social planner's problem. In '*data-heavy*' platforms, VOTING and SOCIAL are the same.

Panel B ranks the total output in four cases. In '*data-light*' platforms, the centralized platform has the largest output. Two decentralized regimes have the same total output while social planner ranks lowest. In '*data-heavy*' platforms, voting-only and social planner have the same total output.

Panel C explicitly presents the total output of the centrally governed platform founder versus buy-back and voting regime in '*data-heavy*' platforms. When A_0 is around 1.5, there exists someplace where the buy-back and the voting regime have a higher or equal output than the centralized platform.

Proposition 5 can also be considered as the reason for the emergence of the decentralized platform. Since the buy-back and the voting regime is able to attract more initial users to the platform, and therefore increase first-period user adoption. It is possible that the buy-back and voting generate more total output than the centralized platform. As a platform founder, this provides an incentive for the platform founder to establish a decentralized platform with a buy-back market.

D. Total user surplus ranking

In this section, we compared the total user surplus in all the cases. The following proposition illustrates the key result:

Proposition 6 Fairness: Comparison on total user surplus. Given the same A_0, the total user surplus ($S = S_1 + \beta\, S_2$) of the voting-only regime (VOTING), the buy-back and voting regime (BUY-BACK), the centralized platform (CENTRAL), and the social planner (SOCIAL) have the following rank:

$$S_{SOCIAL} \geq S_{VOTING} \geq S_{CENTRAL} \tag{21}$$

$$S_{BUY-BACK} <> S_{CENTRAL} \tag{22}$$

Proposition 6 implies that transferring from low-cost users to high-cost users always decreases social welfare. Figure 10 presents the total user surplus ranking ($\overline{A} = 1, \beta = 1$). In this case, '*data-light*' and '*data-heavy*' platforms are separated by a threshold $A_0 = 1$.

Panel A is the total user surplus of benchmarks and voting-only as a function of A_0. Similar to figure 9, the voting-only regime is in the middle of a centralized platform owner and social planner in '*data-light*' platforms. In '*data-heavy*' platforms, voting-only is the same as the social planner's problem.

Panel B ranks the total user surplus in four cases. In '*data-light*' platforms, same as 9, the buy-back market collapses and two decentralized regimes are the same. The centrally governed platform is the lowest one and the social planner is the highest one. In '*data-heavy*' platforms, voting-only and social planner have the same total user surplus.

Panel C explicitly presents the total user surplus of the centrally governed platform founder versus buy-back and voting regime when $A_0 > 1$. Similar to figure 9, when A_0 is around 1.5, there exists someplace where the buy-back and the voting regime have even lower user surplus than the centrally governed platform.

Similar to Section V.C, when A_0 is extremely large, all three cases will have zero social welfare.

E. Policy intervention that limits the extent of data sharing

In the previous section, we require that $\overline{A} \geq A_1 \geq A_0$ and $\overline{A} > 1$. In this section, we discuss a lower $\overline{A}$ case, that is $0 \leq A_0 \leq A_1 \leq \overline{A} \leq 1$. A simple explanation for this assumption is that the government wants to impose restrictions on data sharing. Therefore, by law, the maximum amount of data that can be shared is no larger than the low-cost type. We have the following lemma.

Lemma 7: Properties of A_θ^* when $\overline{A} \leq 1$. (1) Optimal A_θ^* for $\theta \in [0, 1]$ is still continuous and non-decreasing. (2) Optimal A_θ^* for $\theta \in [0, 1]$ is a non-decreasing function of $\overline{A}$.

Therefore, the key difference for a lower $\overline{A}$ is that, pivotal voters optimal $A^*_{\theta_{pivotal}}$ may also decrease and lead to a non-increasing A_1^*.

Proposition 7: Properties of A_1^* and n_2^* with $\overline{A} \leq 1$.(1) A_1^* and n_2^* are non-decreasing with the rise of $\overline{A}$. (2) Second-period total output is non-decreasing in $\overline{A}$.

Proposition 7 illustrates that imposing restrictions on data sharing, in general, limits the ability of low-cost users to express their ideas. As a result, both total output and user surplus decrease. However, these restrictions also offer certain benefits.

User adoption equals $n_2 = 1 - \overline{A}$ when $A_1 = \overline{A}$. This indicates the necessity of implementing restrictions on data sharing in a centralized platform economy. A lower value for $\overline{A}$ decreases the efficiency of the transfer policy. When $A_1 = \overline{A}$, the cost for any $\theta < \overline{A}$ becomes infinite, rendering the transfer policy ineffective. Consequently, the centralized platform owner cannot increase her own benefit at the users' cost.

In general, a lower $\overline{A}$ is more efficient in a centralized platform. However, in a decentralized platform, Imposing restrictions can limit the platform's development.

VI. Conclusion

In this paper, we develop a two-period model to study the efficiency and fairness of DAO with governance tokens. The platform founder designs the platform's decentralized governance rules to maximize the total output of the platform's users. The output of the platform depends on the extent of data sharing and user adoption. The platform's output is consumed by users. The tool to achieve decentralization is the governance token. It is freely distributed to users as a reward for participation and is associated with voting rights on an incremental level of data sharing. The platform founder's decision consists of a voting rule threshold, a buy-back decision, and the level of the buy-back price. Two alternative cases are discussed in the paper: voting-only, buy-back and voting. Heterogenous users have different costs for sharing data due to privacy concerns. They also make decisions on whether to participate in the platform in both periods. Higher data requirement enhances output while increasing user costs.

Our study offers two key insights into tokenomics and decentralized platforms. Firstly, we demonstrate that a decentralized platform with a voting-only regime can yield a platform output that is slightly higher than that of a social planner solution, yet always inferior to the output generated by a centralized platform. Conversely, the user surplus follows an opposite pattern. Thus, decentralized governance through voting trades off user surplus and platform output from the benchmark governance regimes. Secondly, in '*data-heavy*' platforms we find that adopting a token buy-back mechanism as a decentralized governance tool can sometimes result in an equal or higher platform output compared to the centralized platform. This highlights the reason for a platform founder to set up a platform with decentralized token-based governance, rather than a traditional corporate governance structure. It also underscores the advantage of issuing tradable tokens and freely distributing them to users.

Our results hold when data is as significant as real capital in the platform's production function. Additionally, when data and real capital have identical elasticity with respect to the platform's output, utilizing the "buy-back and voting regime" can always achieve a weakly higher total output than the "voting-only regime". This is due to the platform founder being able to commit to a transfer mechanism by allowing token buy-back, which induces early adoption of the platform. This further suggests that issuing tradeable tokens is a part of the optimal choice in decentralized governance.

REFERENCES

Abowd, John M, and Ian M Schmutte, 2019, An economic analysis of privacy protection and statistical accuracy as social choices, *American Economic Review* 109, 171–202.

Acquisti, Alessandro, Curtis Taylor, and Liad Wagman, 2016, The economics of privacy, *Journal of economic Literature* 54, 442–492.

Alchian, Armen A, and Harold Demsetz, 1972, Production, information costs, and economic organization, *The American Economic Review* 62, 777–795.

Appel, Ian, and Jillian Grennan, 2023a, Control of decentralized autonomous organizations, *AEA Papers and Proceedings* 113, 182–85.

Appel, Ian, and Jillian Grennan, 2023b, Decentralized governance and digital asset prices, *Available at SSRN 4367209* .

Arrieta-Ibarra, Imanol, Leonard Goff, Diego Jiménez-Hernández, Jaron Lanier, and E Glen Weyl, 2018, Should we treat data as labor? moving beyond "free", in *aea Papers and Proceedings*, volume 108, 38–42, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Arruñada, Benito, and Luis Garicano, 2018, Blockchain: The birth of decentralized governance, *Pompeu Fabra University, Economics and Business Working Paper Series* 1608.

Bakos, Yannis, and Hanna Halaburda, 2022, Overcoming the coordination problem in new marketplaces via cryptographic tokens, *Available at SSRN 4028849* .

Benhaim, Alon, Brett Hemenway Falk, and Gerry Tsoukalas, 2023a, Balancing power in decentralized governance: Quadratic voting under imperfect information, *Available at SSRN* .

Benhaim, Alon, Brett Hemenway Falk, and Gerry Tsoukalas, 2023b, Scaling blockchains: Can committee-based consensus help?, *Available at SSRN 3914471* .

Bessen, James, 2022, *The New Goliaths: How Corporations Use Software to Dominate Industries, Kill Innovation, and Undermine Regulation* (Yale University Press).

Broccardo, Eleonora, Oliver Hart, and Luigi Zingales, 2020, Exit vs. voice, *Journal of Political Economy* .

Brunnermeier, Markus K, Sebastian A Merkel, and Yuliy Sannikov, 2022, Safe assets: A dynamic retrading perspective, *Journal of Political Economy* Forthcoming.

Cachon, Gerard P, Tolga Dizdarer, and Gerry Tsoukalas, 2022, Pricing control and regulation on online service platforms, *Available at SSRN 3957209* .

Cong, Lin William, Ye Li, and Neng Wang, 2021a, Tokenomics: Dynamic Adoption and Valuation, *The Review of Financial Studies* 34, 1105–1155.

Cong, Lin William, Ye Li, and Neng Wang, 2022a, Token-based platform finance, *Journal of Financial Economics* 144, 972–991.

Cong, Lin William, and Simon Mayer, 2023, Antitrust, regulation, and user union in the era of digital platforms and big data, *NBER working paper w30881* .

Cong, Lin William, Wenshi Wei, Danxia Xie, and Longtian Zhang, 2022b, Endogenous growth under multiple uses of data, *Journal of Economic Dynamics and Control* 141, 104395.

Cong, Lin William, Danxia Xie, and Longtian Zhang, 2021b, Knowledge accumulation, privacy, and growth in a data economy, *Management Science* 67, 6480–6492.

Fainmesser, Itay P, Andrea Galeotti, and Ruslan Momot, 2023, Digital privacy, *Management Science* 69, 3157–3173.

Farboodi, Maryam, and Laura Veldkamp, 2022, A model of the data economy, *NBER working paper w28427* .

Galperti, Simone, and Jacopo Perego, 2022, Competitive markets for personal data, *Available at SSRN 4309966* .

Gan, Jingxing, Gerry Tsoukalas, and Serguei Netessine, 2021, Initial coin offerings, speculation, and asset tokenization, *Management Science* 67, 914–931.

Goldfarb, Avi, and Catherine Tucker, 2012, Shifts in privacy concerns, *American Economic Review* 102, 349–353.

Goldfarb, Avi, and Catherine E Tucker, 2011, Privacy regulation and online advertising, *Management science* 57, 57–71.

Han, Jungsuk, Jongsub Lee, and Tao Li, 2023, Dao governance, *Available at SSRN 4028849* .

Hansmann, Henry, 2000, *The Ownership of Enterprise* (Harvard University Press).

Ichihashi, Shota, 2021, Competing data intermediaries, *The RAND Journal of Economics* 52, 515–537.

Jones, Charles I, and Christopher Tonetti, 2020, Nonrivalry and the Economics of Data, *American Economic Review* 110, 2819–2858.

Kummer, Michael, and Patrick Schulte, 2019, When private information settles the bill: Money and privacy in google's market for smartphone applications, *Management Science* 65, 3470–3494.

Li, Jiasun, and William Mann, 2018, Digital tokens and platform building .

Lin, Tesary, 2022, Valuing intrinsic and instrumental preferences for privacy, *Marketing Science* 41, 663–681.

Lommers, Kristof, Jiahua Xu, and Teng Andrea Xu, 2022, A framework for dao token valuation, *Available at SSRN 4309744* .

Mayer, Simon, 2019, Token-Based Platforms and Speculators, *SSRN Electronic Journal* .

Miller, Amalia R, and Catherine Tucker, 2009, Privacy protection and technology diffusion: The case of electronic medical records, *Management science* 55, 1077–1093.

Miller, Amalia R, and Catherine Tucker, 2018, Privacy protection, personalized medicine, and genetic testing, *Management Science* 64, 4648–4668.

Quan, Junjun, 2022, Tracing out international data flow: The value of data and privacy, *Available at SSRN 4185169* .

Ramirez, Edith, Julie Brill, Maureen K. Ohlhausen, Joshua D. Wright, and Terrell McSweeny, 2014, *Data Brokers: A Call For Transparency and Accountability: A Report of the Federal Trade Commission* (Federal Trade Commission).

Romer, Paul M, 1990, Endogenous technological change, *Journal of political Economy* 98, S71–S102.

Sockin, Michael, and Wei Xiong, 2023a, Decentralization through tokenization, *The Journal of Finance* 78, 247–299.

Sockin, Michael, and Wei Xiong, 2023b, A model of cryptocurrencies, *Management Science* .

Tang, Huan, 2019, The value of privacy: Evidence from online borrowers, *Available at SSRN* .

Tsoukalas, Gerry, and Brett Hemenway Falk, 2020, Token-weighted crowdsourcing, *Management Science* 66, 3843–3859.

Veldkamp, Laura, and Cindy Chung, 2023, Data and the aggregate economy, *Journal of Economic Literature* .

Appendix

A. Proofs

Proof of Lemma 1

The cost function is $C_t(A_{t-1};\theta) = \mathbb{1}\{\theta < A_{t-1}\}\frac{A_{t-1}-\theta}{\overline{A}-A_{t-1}}$, while $\theta \in [0,1]$, $A_{t-1} \in (0,\overline{A}]$, and exogenous $\overline{A} > 1$.

(1) To prove $C_t(A_{t-1};\theta)$ is a weakly increasing function of A_{t-1} for $\forall\theta$, we consider two cases. First, for $\forall\theta$, when $A_{t-1} \leq \theta$, the cost function is constant zero. Also, $C_t(A_{t-1};\theta)$ is continuous at $A_{t-1} = \theta$.

Second, when $A_{t-1} > \theta$, the first order derivative of cost function with respect to A_{t-1} is:

$$\frac{\partial C_t(A_{t-1};\theta)}{\partial A_{t-1}} = \frac{\overline{A}-\theta}{(\overline{A}-A_{t-1})^2}$$

Since $\overline{A} > 1 \geq \theta$ always exists, we have $\frac{\partial C_t(A_{t-1};\theta)}{\partial A_{t-1}} > 0$ when $A_{t-1} > \theta$.

(2) To prove $C_t(A_{t-1};\theta)$ is a weakly decreasing function of θ for $\forall A_{t-1}$, we want to show that for any $1 \geq \theta_H > \theta_L \geq 0$, we have $C_t(A_{t-1};\theta_H) = \mathbb{1}\{\theta_H < A_{t-1}\}\frac{A_{t-1}-\theta_H}{\overline{A}-A_{t-1}} \geq C_t(A_{t-1};\theta_L) = \mathbb{1}\{\theta_L < A_{t-1}\}\frac{A_{t-1}-\theta_L}{\overline{A}-A_{t-1}}$.

Given A_{t-1}, there are three possible cases. First, when $A_{t-1} > \theta_H > \theta_L$, $C_t(A_{t-1};\theta_H) = \frac{A_{t-1}-\theta_H}{\overline{A}-A_{t-1}} < C_t(A_{t-1};\theta_L) = \frac{A_{t-1}-\theta_L}{\overline{A}-A_{t-1}}$. Second, if $\theta_H \geq A_{t-1} > \theta_L$, $C_t(A_{t-1};\theta_H) = 0 < C_t(A_{t-1};\theta_L) = \frac{A_{t-1}-\theta_L}{\overline{A}-A_{t-1}}$. Third, if $\theta_H > \theta_L \geq A_{t-1}$, we have $C_t(A_{t-1};\theta_H) = C_t(A_{t-1};\theta_L) = 0$.

(3) When $\theta < A_{t-1}$, the derivatives of cost function $C_t(A_{t-1};\theta)$ with respect to A_{t-1} is $\frac{\partial C_t(A_{t-1};\theta)}{\partial A_{t-1}} = \frac{\overline{A}-\theta}{(\overline{A}-A_{t-1})^2}$. Clearly, $\frac{\partial C_t(A_{t-1};\theta)}{\partial A_{t-1}}$ is a decreasing function of θ.

Proof of Lemma 2

The utility function is $U_t(A_{t-1};\theta) = c_t - C_t(A_{t-1};\theta) = A_{t-1} - \mathbb{1}\{\theta < A_{t-1}\}\frac{A_{t-1}-\theta}{\overline{A}-A_{t-1}}$ while $\theta \in [0,1]$, $A_{t-1} \in (0,\overline{A}]$, and exogenous $\overline{A} > 1$.

(1) To prove $U_t(A_{t-1};\theta)$ is a weakly increasing function of θ for $\forall A_{t-1}$, we consider two cases.

First, when $\theta \geq A_{t-1}$, $U_t(A_{t-1};\theta) = A_{t-1}$ is constant for $\forall\theta$. Second, when $\theta < A_{t-1}$, $U_t(A_{t-1};\theta) = A_{t-1} - \frac{A_{t-1}-\theta}{\overline{A}-A_{t-1}}$ is a strictly increasing function of θ for $\forall A_{t-1}$.

(2) When $A_L < \theta$, $U_t(A_L;\theta) = A_L > 0$. It is easy to see that, $U_t(A_L;\theta) \leq 0$ occurs when $A_L \geq \theta$. Therefore, we focus on discussing the properties of the utility function $U_t(A_{t-1};\theta) = A_{t-1} - \frac{A_{t-1}-\theta}{\overline{A}-A_{t-1}}$ which appears when $A_{t-1} \geq \theta$.

The first order derivative of the utility function with respect to A_{t-1} is:

$$\frac{\partial U_t(A_{t-1};\theta)}{\partial A_{t-1}} = 1 - \frac{\overline{A}-\theta}{(\overline{A}-A_{t-1})^2}$$

It is a strictly decreasing function of A_{t-1}. Therefore, $\frac{\partial U_t(A_{t-1};\theta)}{\partial A_{t-1}}$ achieves its max at $A_{t-1} = \theta$, which is:

$$\frac{\partial U_t(A_{t-1};\theta)}{\partial A_{t-1}}|_{A_{t-1}=\theta} = 1 - \frac{1}{\overline{A}-\theta}$$

There are two possible cases depending on the value of $\overline{A}$. First, if $\overline{A}-\theta < 1$, $\frac{\partial U_t(A_{t-1};\theta)}{\partial A_{t-1}} < 0$ for $\forall A_{t-1} \geq \theta$, which means that $U_t(A_{t-1};\theta)$ is strictly decreasing function of A_{t-1}. Therefore, this lemma always holds as long as $A_H > A_L \geq \theta$.

Second, if $\overline{A}-\theta \geq 1$, $\frac{\partial U_t(A_{t-1};\theta)}{\partial A_{t-1}}$ is larger than 0 when $A_{t-1}=\theta$. Then, with the rise of A_{t-1}, $\frac{\partial U_t(A_{t-1};\theta)}{\partial A_{t-1}}$ gradually becomes small than zero. For example, when A_{t-1} approaches $\overline{A}$, $\frac{\partial U_t(A_{t-1};\theta)}{\partial A_{t-1}}$ goes to negative infinity. This implies when $A_{t-1} \in [\theta, \overline{A}]$, $\frac{\partial U_t(A_{t-1};\theta)}{\partial A_{t-1}}$ starts with a positive number and then turns to negative. Therefore, the utility function $U_t(A_{t-1};\theta)$ first increases and then decreases when $A_{t-1} \geq \theta$. Similarly, the utility function $U_t(A_{t-1};\theta) = A_{t-1} - \frac{A_{t-1}-\theta}{\overline{A}-A_{t-1}}$ is large than zero when $A_{t-1}=\theta$. $U_t(A_{t-1};\theta)$ first increases and then decreases to gradually become smaller than zero when A_{t-1} approaches $\overline{A}$. Since we restrict $U_t(A_L;\theta) \leq 0$, it indicates that $\frac{\partial U_t(A_{t-1};\theta)}{\partial A_{t-1}}|_{A_{t-1}=\theta_L} < 0$. Also, for any $A_{t-1} \geq A_L$, we always have $\frac{\partial U_t(A_{t-1};\theta)}{\partial A_{t-1}} < 0$. Thus, for any $A_H \geq A_L$, $U_t(A_H;\theta) \leq U_t(A_L;\theta) \leq 0$ always hold.

Proof of Lemma 3

The user's optimal choice of second-period data-sharing A^*_θ is coming from the following maximization problem:

$$A^*_\theta = argmax_{A_1 \in [A_0,\overline{A}]}\ A_1 - \mathbb{1}\{\theta < A_1\}\frac{A_1-\theta}{\overline{A}-A_1}$$

(1) It is easy to see that the maximization of $A_1 - \mathbb{1}\{\theta < A_1\}\frac{A_1-\theta}{\overline{A}-A_1}$ is achieved when $\theta < A_1 < \overline{A}$. When $A_1 \leq \theta$, the utility function is strictly increasing and achieves its maximum at $A_1 = \theta$.

As discussed in Lemma 2 (2), there are two possible cases for $\frac{\partial U_t(A_{t-1};\theta)}{\partial A_{t-1}}$ depending on the value of $\overline{A}$.

First, if $\overline{A}-\theta < 1$, $\frac{\partial U_t(A_{t-1};\theta)}{\partial A_{t-1}} < 0$ when $A_1 \geq \theta$. In this case $A^*_\theta = \theta$ since the $A_1 - \frac{A_1-\theta}{\overline{A}-A_1}$ is decreasing with the rise of A_1. Second, if $\overline{A}-\theta \geq 1$, $\frac{\partial U_t(A_{t-1};\theta)}{\partial A_{t-1}}$ starts with positive and then turns to negative. It implies that $U_t(A_{t-1};\theta)$ first increases and then decreases with A_1 and $A^*_\theta > \theta$.

Combining the two cases together, we will have $A^*_\theta \geq \theta$.

(2) Similar to the discussion in Lemma 3 (1), when $\overline{A}-1 \geq \theta$, the first order derivative of the $A_1 - \frac{A_1-\theta}{\overline{A}-A_1}$ with respect to A_1 is $1 - \frac{\overline{A}-\theta}{(\overline{A}-A_1)^2}$. Therefore, the optimal $A^*_\theta = \overline{A} - \sqrt{\overline{A}-\theta}$. It is an increasing function of θ. Specifically, when $\theta = \overline{A}-1$, $A^*_\theta = \overline{A}-1 = \theta$. When $\overline{A}-1 < \theta$ and correspondingly $\frac{\partial U_t(A_{t-1};\theta)}{\partial A_{t-1}} < 0$, $A^*_\theta = \theta$ is an increasing function of θ.

(3) According to Lemma 3 (2), for $\forall \theta_H > \theta_L$, we have $A^*_{\theta_H} > A^*_{\theta_L}$. Therefore, due to the optimality of $A^*_{\theta_H}$, we have $U_t(A^*_{\theta_H};\theta_H) \geq U_t(A^*_{\theta_L};\theta_H)$. According to Lemma 2 (1), we have $U_t(A^*_{\theta_L};\theta_H) \geq U_t(A^*_{\theta_L};\theta_L)$. Finally, we have $U_t(A^*_{\theta_H};\theta_H) \geq U_t(A^*_{\theta_L};\theta_L)$ which implies that $U^*_{2,\theta}$ is weakly increasing in θ.

Proof of Proposition 1

Due to the voting mechanism's design, when a proposal a fails, the platform introduces a new proposal with $a - \epsilon$, where $\epsilon > 0$ is infinitely small. This process is repeated until a proposal is approved.

Consequently, for any user with type θ and the corresponding optimal A^*_θ, his utility is maximized at A^*_θ and diminishes if the final A_1 deviates from A^*_θ.

When confronted with a proposal, a user faces a binary choice: 'yes' or 'no'. The user selects the option that yields the highest expected utility. If the proposal raises an incremental data-sharing requirement $a > A^*_\theta - A_0$, the user can consistently anticipate that the next proposal will increase his utility. Voting 'no' in this case would raise the probability of the current proposal's failure, while voting 'yes' would boost the chances of approval for this proposal. Therefore, the user will vote for 'no' and wait for the next proposal. However, if $a \leq A^*_\theta - A_0$, the user will anticipate that the next proposal will reduce his utility. Since the user aims to prevent the failure of the proposal, he will vote 'yes'.

Proof of Lemma 4

First, without any restrictions, we want to find $A^*_1 = A_{max}$ and corresponding $n_2(A_{max}) = max\{0, A_{max} - A_{max}(\overline{A} - A_{max})\}$ that maximizes second-period total output $A_{max}n_2$. In this case, we solve the question below:

$$A_{max} = argmax_{A_1 \in (0, A^*_{\theta=1}]}\, A_1[1 - max\{0, A_1 - A_1(\overline{A} - A_1)\}]$$

where $A^*_{\theta=1}$ is the optimal second-period data sharing for the lowest-cost type.

We consider two cases. The first case is when $\overline{A} - 1 > 1$ and therefore $A^*_{\theta=1} = \overline{A} - \sqrt{\overline{A} - 1} < \overline{A} - 1$. Consequently, we have $A_1 \leq \overline{A} - 1$ and $max\{0, A_1 - A_1(\overline{A} - A_1)\} = 0$. In this case, the A_1 should be the larger the better. Therefore $A_{max} = \overline{A} - \sqrt{\overline{A} - 1}$.

The second case is when $\overline{A} - 1 \leq 1$ and therefore $A^*_{\theta=1} = \overline{A} - \sqrt{\overline{A} - 1} \geq \overline{A} - 1$. In this case, there are two possible solutions. If $A_1 \leq \overline{A} - 1$, we have $A_{max} = \overline{A} - 1$ and $n_2 = 1$. If $A_1 > \overline{A} - 1$, solving the maximization problem below, we have the $A_{max} = \frac{\overline{A}-1+\sqrt{(\overline{A}-1)^2+3}}{3}$ and $n_2 = [1 - (\frac{\overline{A}-1+\sqrt{(\overline{A}-1)^2+3}}{3} - \frac{\overline{A}-1+\sqrt{(\overline{A}-1)^2+3}}{3}(\overline{A} - \frac{\overline{A}-1+\sqrt{(\overline{A}-1)^2+3}}{3})]$:

$$A_{max} = argmax\, A_1[1 - (A_1 - A_1(\overline{A} - A_1))]$$

One can prove that $\frac{\overline{A}-1+\sqrt{(\overline{A}-1)^2+3}}{3} \in [\overline{A} - 1, \overline{A} - \sqrt{\overline{A} - 1}]$ when $\overline{A} - 1 \leq 1$. Also, when $A_{max} = \frac{\overline{A}-1+\sqrt{(\overline{A}-1)^2+3}}{3}$, $A_{max}n_2(A_{max})$ is a decreasing function of $\overline{A}$. Therefore, when $\overline{A} < \overline{A}_{root}$, $A_{max} = \frac{\overline{A}-1+\sqrt{(\overline{A}-1)^2+3}}{3}$. When $\overline{A} \geq \overline{A}_{root}$, $A_{max} = \overline{A} - 1$. Denote $A_{max1} = \frac{\overline{A}-1+\sqrt{(\overline{A}-1)^2+3}}{3}$, the root $\overline{A}_{root} \approx 1.238$ is obtained below:

$$A_{max1}[1 - (A_{max1} - A_{max1}(\overline{A} - A_{max1}))] = \overline{A} - 1$$

Therefore, if $A_0 \leq A_{max}$ (in the 'data-light'), which means A_{max} is achievable using τ or/and p. Therefore, the platform founder will develop the mechanism such that the voting outcome $A_1^* = A_{max}$. If $A_0 > A_{max}$ (in the 'data-heavy') platform the A_{max} is not achievable. More deviation from A_{max} means a lower output. Therefore, platform founder will design τ and/or p such that $A_1^* = A_0$.

Proof of Proposition 2

Before buy-back, users with type $[\theta_1, 1]$ have governance tokens and decide whether they should sell governance tokens and leave the platform or stay in the platform to produce.

If $p = 0$, this question falls back to the voting-only regime, and there is zero supply for the governance token and no market for buy-back. If $p \geq U_2(A^*_{\theta=1}; \theta = 1)$, which is the maximum utility type $\theta = 1$ could achieve, then everyone wants to sell the governance token. When p increases, the willingness to sell the governance token also increases. Therefore, there exists some p in between such that the supply and demand are non-zero.

θ_{sell} is the most important type such that his second-period utility to sell the governance token is equal to staying in the platform. For any users with type $\theta_{sell} \geq \theta \geq \theta_1$, they choose to sell the governance token and leave. They are not allowed to participate in second-period production and their second-period utility is p.

Proof of Proposition 3

For buy-back and voting regime, we allow $p \geq 0$. When $p = 0$, the buy-back regime collapses to the voting-only regime. Under the BUY-BACK, it is always possible that the platform founder chooses $p = 0$ if she finds this creates a larger total output. In other words, at buy-back optimal, any $p > 0$ will generate higher total output for the platform founder compared to the voting-only regime. This implies that the voting-only regime is a special case of buy-back and voting regime.

Proof of Proposition 4

The second-period total output equals $Y_2 = A_1 n_2$. Given A_1, if there is no interfere, $n_2 = 1 - \theta_2$ is decided by $U_t(A_1; \theta_2) = 0$. Also, $S_2 > 0$ if $n_2 > 0$.

Fixed the second-period data-sharing requirement A_1, the centralized platform owner carries the following policy: transfer the final goods from low-cost users to potential high-cost users who will not join the platform if there is no subsidy. On the other hand, the centralized platform owner guaranteed that everyone on this platform has a non-negative second-period utility (by at most taking $U_2(A_1; \theta)$ from type θ users). Therefore, no one has the incentive to leave the platform but this transfer policy increases n_2 and the second-period total output.

Proof of Lemma 5

As discussed in proposition 4, given A_1, the centralized platform owner will attract as many users as she can to increase the second-period total output. The maximum resource she can use is:

$$\int_{\underline{\theta}}^{1} U_2(\theta)d\theta$$

where $\underline{\theta}$ is the lowest type that will join the platform in the second period with data-sharing requirement A_1 if there is no interference.

Therefore, if the centralized platform owner uses all resources to subsidize, she can attract additional $\underline{\theta} - \theta_{root}$ users into the platform. θ_{root} is obtained by:

$$-\int_{\theta_{root}}^{\underline{\theta}} U_2(\theta)d\theta = \int_{\underline{\theta}}^{1} U_2(\theta)d\theta$$

Simplifying the expression, we can get the θ_{root} from the equation $\int_{\theta_{root}}^{1} U_2(\theta)d\theta = 0$, where $U_2 = A_1 - \mathbb{1}\{\theta < A_1\}\frac{A_1-\theta}{\overline{A}-A_1}$.

It is possible that $\theta_{root} < 0$, thus the lowest type that participates in second-period production is $\theta_2 = max\{0, \theta_{root}\}$. The second-period user adoption n_2 equals $1 - \theta_2$. If $\theta_2 > 0$, this implies that the centralized platform owner has already used all the resources collected from low-cost users. However, she is still not able to attract any users from $[0, \theta_2)$ to join the platform. Consequently, the second-period user surplus is zero.

Proof of Lemma 6

Given A_{t-1}, the social planner chooses θ_t to maximize the user surplus. user surplus can be expressed as:

$$\int_{\theta_t}^{1} U_t(\theta)d\theta = A_{t-1}(1-\theta_t) + \frac{1}{2(\overline{A}-A_{t-1})}[\mathbb{1}\{A_{t-1} > 1\}(A_{t-1}-1)^2 - \mathbb{1}\{A_{t-1} > \theta_t\}(A_{t-1}-\theta_t)^2]$$

The first order condition is:

$$-A_{t-1} + \frac{1}{(\overline{A}-A_{t-1})}[\mathbb{1}\{A_{t-1} > \theta_t\}(A_{t-1}-\theta_t)]$$

Also, one may notice that the optimal θ_t should be smaller than A_{t-1}. Since for $U_t = A_{t-1} - \mathbb{1}\{\theta < A_{t-1}\}\frac{A_{t-1}-\theta}{\overline{A}-A_{t-1}}$, there $\exists \theta < A_{t-1}$ such that $U_t(A_{t-1};\theta) > 0$. Including this user in the production process will increase the total user surplus. Therefore, solving the first order condition, the optimal θ_t^* is equal to $\theta_t^* = max\{0, A_{t-1} - A_{t-1}(\overline{A} - A_{t-1})\}$.

Proof of Proposition 5

By definition, total output Y equals to:

$$Y_1 + \beta Y_2 = A_0 n_1 + \beta A_1 n_2 \tag{23}$$

In the first period, A_0 is exogenous, and n_1 is the same for the social planner, voting, and central problem since the governance starts at $t = 1$. Therefore, to prove equation (16), we only need to verify:

$$Y_2^{SOCIAL} \leq Y_2^{VOTING} \leq Y_2^{CENTRAL} \tag{24}$$

The right part of the inequality $Y_2^{VOTING} \leq Y_2^{CENTRAL}$ is easy to prove. Since for any A_1, the centralized platform owner can use transfer policy to attract more users into the platform and make sure $n_2^C \geq n_2^{VOTING}$. At the same time, the centralized platform owner chases high total output for any A_1. Therefore, $Y_2^{VOTING} \leq Y_2^{CENTRAL}$ always holds.

Also, according to Lemma 5, for the same A_1, the social planner and voting-only regime will have the same n_2 since the social planner will not allow negative utility users to enter the platform. For the voting-only regime, with different τ, the range of $A_1 \in [A_0, A^*_{\theta=1}]$ is predetermined. $A^*_{\theta=1} = \overline{A} - \sqrt{\overline{A} - 1}$ is the optimal second-period data sharing for the lowest-cost type. Social planner is able to choose any $A_1 \in [A_0, \overline{A}]$. However, when $A_1 \in (A^*_{\theta=1}, \overline{A}]$, a little increase from $A^*_{\theta=1}$ will harm all users' utility thus decrease the overall welfare. Therefore, for both social planner and voting-only regimes, they are choosing within the same pairs (A_1, n_2) to maximize their goal.

Therefore, considering the optimal voting-only regime, with $A_1^{*VOTING}$ and $n_2^{*VOTING}$, social planners still have an incentive to move away from this output to increase social welfare. Therefore, we have $Y_{SOCIAL} \leq Y_{VOTING}$.

The proof of equation (17) is a combination of the proof of equation (16) and proposition 3.

The proof of inequality (18) is in figure 9, where the total output of central exceeds buy-back with a lower A_0 and gradually becomes equal or even smaller when A_0 is around 1.5.

Proof of Proposition 6

By definition, total user surplus S equals to:

$$S_1 + \beta\, S_2 = \int_{\theta_1}^{1} U_1 d\theta + \beta \int_{\theta_2}^{1} U_2 d\theta \tag{25}$$

In the first period, A_0 is exogenous, and n_1 is the same for the social planner, voting-only, and centralized platform. Therefore, to prove equation (19), we only need to verify:

$$S_2^{SOCIAL} \geq S_2^{VOTING} \geq S_2^{CENTRAL} \tag{26}$$

According to Proposition 5, both social planner and voting-only regimes are choosing within the same pairs (A_1, n_2) to maximize their goal. Since the social planner aims to maximize total user surplus, while decentralized platform founder's goal is to maximize total output. Therefore, with the social planner's $A_1^{*SOCIAL}$ and $n_2^{*SOCIAL}$, the decentralized platform founder still has an incentive to move away from this output to increase total output. We can easily get $S_2^{SOCIAL} \geq Y_2^{VOTING}$.

We will then prove $S_2^{VOTING} \geq S_2^{CENTRAL}$ by contradiction. Suppose at optimal, we have some $S_2^{CENTRAL} > S_2^{VOTING} \geq 0$, this suggest that $\theta_2^{CENTRAL} = 1 - n_2^{CENTRAL} = 0$. Otherwise, the centralized platform owner can attract more users using subsidies. However, an optimal $Y_2^{CENTRAL}$ also suggest that for $\forall A_1 > A_1^{CENTRAL}$, $\theta_2 > 0$. This can guarantee that $Y_2^{CENTRAL}$ is the largest one among all possible pairs (A_1, n_2). A little increase $A_1^{CENTRAL}$ will cause the maximum source that the social planner can use to attract additional investors to decrease and social planner cannot attract all investors into the platform. All the previous argument implies that when at its optimal, $A_1^{CENTRAL}$ is the largest A_1 that makes $\theta_2 = 0$. The social planner just exhausts all her sources to attract the user which further suggests that $S_2^{CENTRAL}$ should be 0. A contradiction is found.

The proof of inequality (20) is in figure 10, where the total user surplus of buy-back exceeds central almost all the time. Only when A_0 is around 1.5, there exist some regions that buy-back has a lower user surplus.

Proof of Lemma 7

For $\forall\theta \leq \overline{A}$, $A^*(\theta) = \overline{A} - \sqrt{\overline{A} - \theta}$. For $\forall\theta > \overline{A}$, $A^*(\theta) = \overline{A}$. Therefore, it is easy to see that A^* for $\theta \in [0,1]$ is still continuous and non-decreasing. Also, optimal A^* for $\theta \in [0,1]$ is a non-decreasing function of $\overline{A}$.

Proof of Proposition 7

When $\overline{A}$ decreases, since $A^*(\theta)$ is non-increasing, therefore, for voting-only and buy-back, $A_1 = max\{A_0, A^*_{\theta_{pivotal}}\}$ is non-increasing. Relatively, for any A_1, with the fall of $\overline{A}$, all users suffer greater cost and n_2 is also non-increasing.

For two benchmarks, for any A_1 they choose, they will receive a corresponding non-increasing n_2. Also, the fall of $\overline{A}$ damages all users and thus decreases overall welfare and induces a non-increasing A_1.

(2) is a direct conclusion from (1).

B. Figures

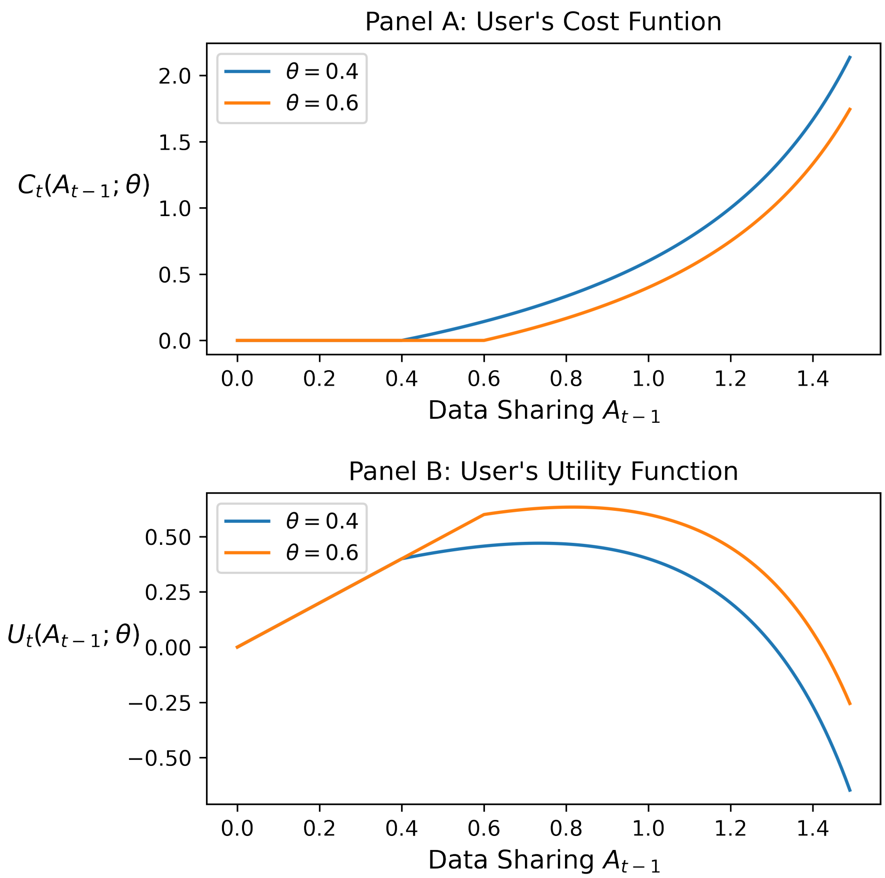

Figure 1. User's cost function and utility function (parameters: $\overline{A} = 2$)

This figure illustrates the varying privacy costs and utility of different users as a function of the data-sharing requirement. Following Lemma 1 and Lemma 2 in Section IV, privacy cost C_t is a weakly increasing function of A_{t-1} for all θ, and similarly, it is a weakly increasing function of θ for all A_{t-1}. User utility exhibits a weakly increasing function of θ for all A_{t-1}. When A_{t-1} becomes large, user utility turns negative, as user's privacy cost starts to dominate the sign.

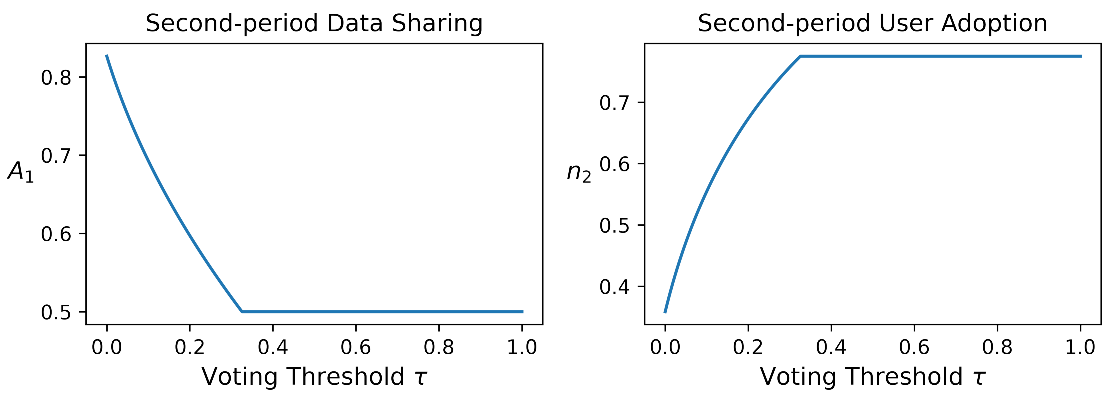

Figure 2. Platform founder's trade-off in designing voting-only governance regime (parameters: $A_0 = 0.5$, $\overline{A} = 1.05$)

This figure shows the primary trade-off faced by a decentralized platform founder when designing a voting-only regime. The platform founder's objective is to maximize the overall output, which is the product of second-period data sharing requirement A_1 and second-period user adoption n_2. As discussed in Section IV.B, A_1 is non-increasing with the rise of the voting threshold τ. Simultaneously, n_2 is affected by A_1 through the output and privacy channel. Consequently, the platform founder chooses between imposing stricter data-sharing requirements or achieving a higher level of user adoption.

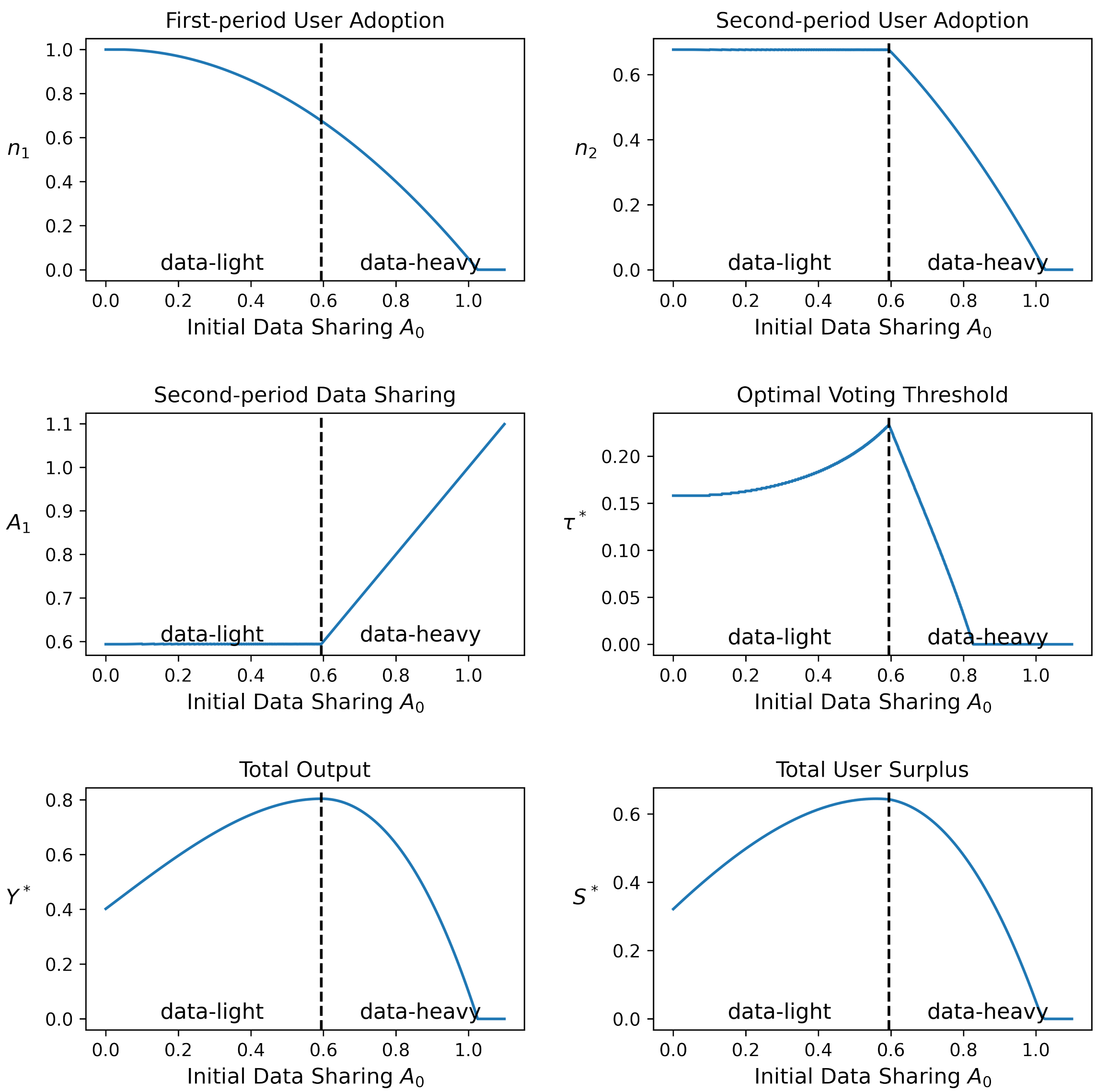

Figure 3. Optimal voting-only governance regime for 'data-light' and 'data-heavy' platforms (parameters: $\beta = 1$, $\overline{A} = 1.05$)

This figure presents the optimal solution for $n_1, n_2, A_1, \tau^*, Y^*, S^*$ as a function of A_0. A_0 is the minimum (initial) requirement on the extent of data sharing which indicates whether the platform's product is 'data-light' or 'data-heavy'. The classification of 'data-light' or 'data-heavy' depends on the $\overline{A}$. In this graph, we label platforms with $A_0 < 0.594$ as 'data-light' platforms, with features that the optimal voting threshold, total output, and total user surplus increase as A_0 grows. Further discussions are presented in Section IV.B.1.

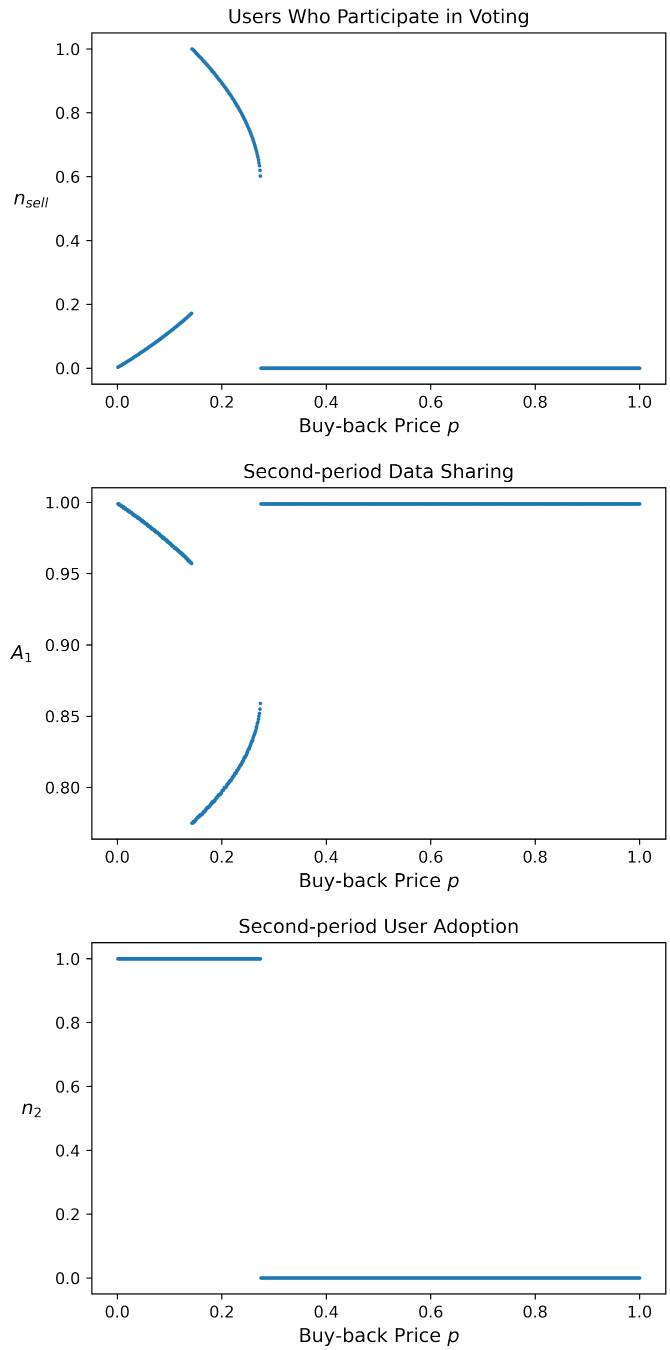

Figure 4. Platform founder's trade-off in designing buy-back and voting governance regime (parameters: $A_0 = 0.2$, $\overline{A} = 2$, $\tau = 0.5$)

This figure illustrates the primary trade-off that a decentralized platform founder faces when designing a buy-back and voting regime. The platform founder's objective is to maximize $Y_2 = n_2 A_1$, where A_1 is influenced by n_{sell}. As discussed in Section IV.C, the impact of the buy-back price p on n_{sell} is complex and depends on initial adoption and selling willingness. The influence of the buy-back price p on A_1 and n_2 is mediated through n_{sell}. When the buy-back price is high, all the token holders want to sell governance tokens and the buy-back market crushes down.

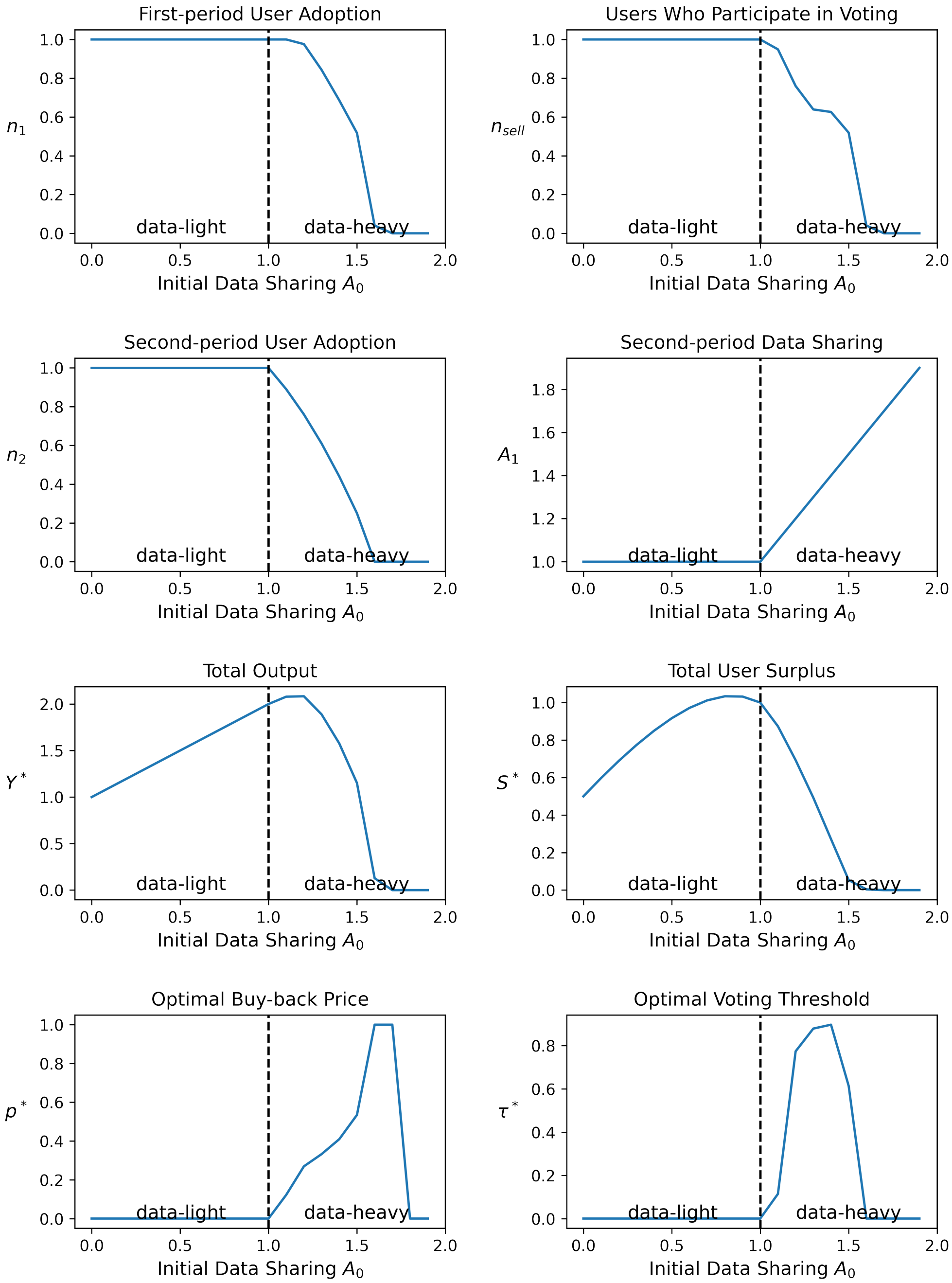

Figure 5. Optimal buy-back and voting governance regime for 'data-light' and 'data-heavy' platforms (parameters: $\overline{A} = 2$)

This figure presents the optimal solution for $n_1, n_{sell}, n_2, A_1, Y^*, S^*, \tau^*, p^*$ as a function of A_0. A_0 is the minimum (initial) requirement on the extent of data sharing which indicates whether the platform's product is 'data-light' or 'data-heavy'. In this graph, we classify platforms with $A_0 < 1$ as 'data-light' platforms, with features that the buy-back market collapses and the buy-back and voting regime becomes a voting-only regime. Further discussions are presented in Section IV.C.1.

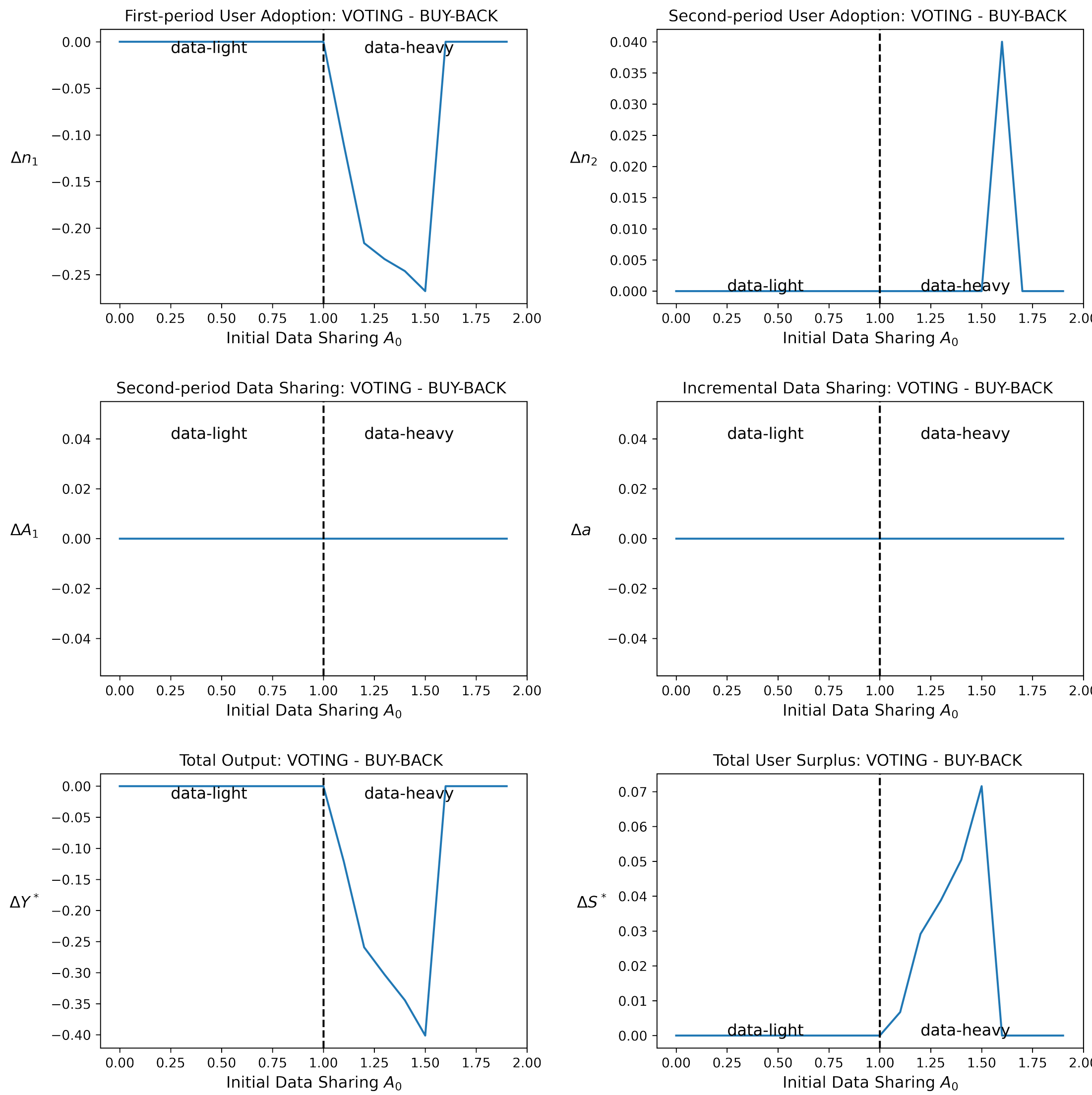

Figure 6. Comparison between the optimal voting-only governance regime and buy-back and voting governance regime (parameters: $\overline{A} = 2$)

This figure presents the difference between the voting-only regime and buy-back and voting regime, which are $\Delta n_1 \equiv n_1^{\text{VOTING}} - n_1^{\text{BUY-BACK}}$, $\Delta n_2 \equiv n_2^{\text{VOTING}} - n_2^{\text{BUY-BACK}}$, $\Delta A_1 \equiv A_1^{\text{VOTING}} - A_1^{\text{BUY-BACK}}$, $\Delta a \equiv a^{\text{VOTING}} - a^{\text{BUY-BACK}}$, $\Delta Y \equiv Y^{\text{VOTING}} - Y^{\text{BUY-BACK}}$, $\Delta S \equiv S^{\text{VOTING}} - S^{\text{BUY-BACK}}$ as a function of A_0. A_0 is the minimum (initial) requirement on the extent of data sharing which indicates whether the platform's product is 'data-light' or 'data-heavy'. In this graph, we denote platforms with $A_0 < 1$ to be 'data-light' platforms. The difference is zero when the platform is 'data-light'. The buy-back market only exists in the 'data-heavy' platforms. Further discussions are presented in Section IV.D.

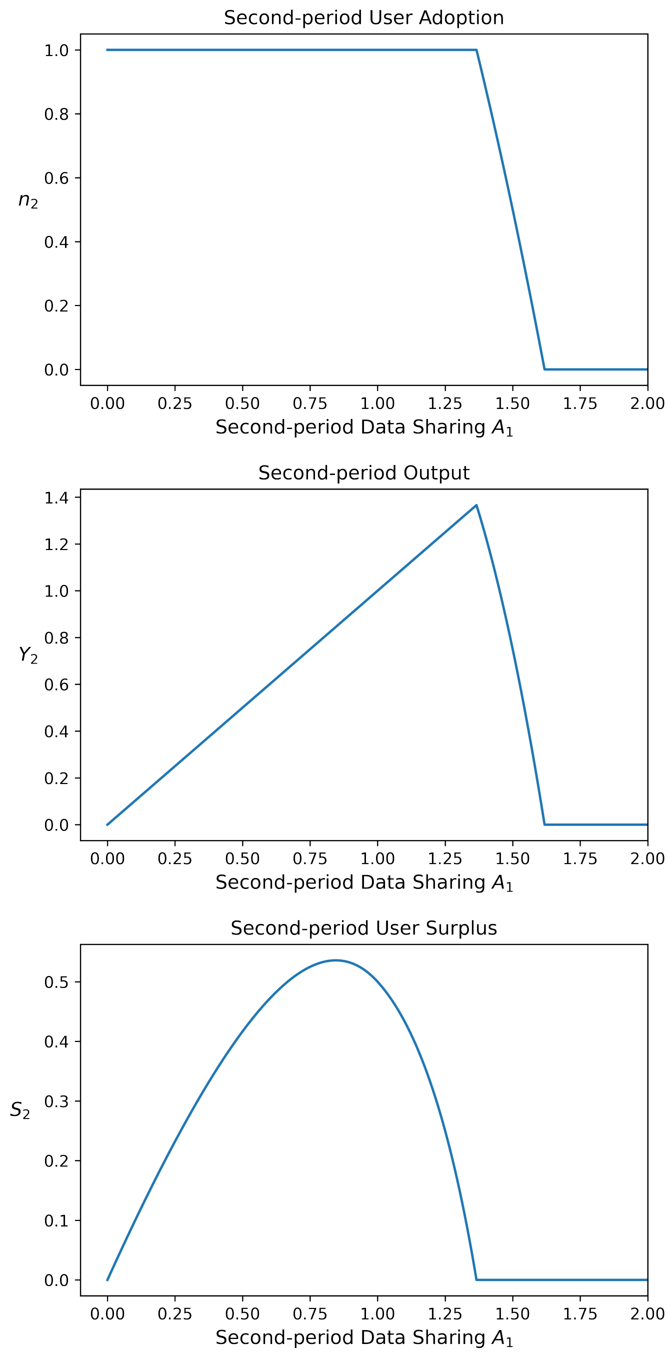

Figure 7. The centralized platform founder's trade-off (parameters: $\overline{A} = 2$)

This figure illustrates the primary trade-off that a centralized platform founder faces when making decisions at $t = 1$. The centralized founder wants to maximize $Y_2 = n_2 A_1$ by choosing A_1 and applying the transfer policy. As discussed in Section V.A, n_2 is non-increasing in A_1. Consequently, the platform founder must make a choice between imposing stricter data-sharing requirements or achieving a higher level of user adoption.

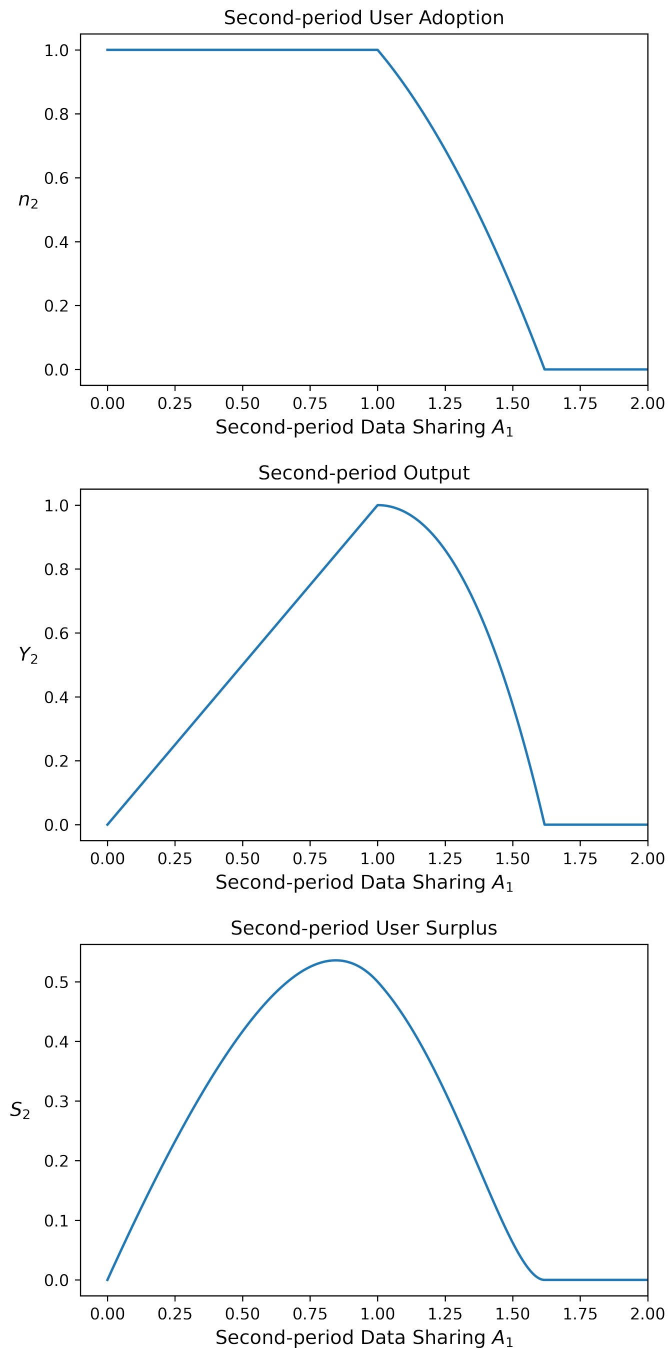

Figure 8. The social planner's trade-off (parameters: $\overline{A} = 2$)

This figure illustrates the primary trade-off of a social planner when making decisions at $t = 1$. The objective of a social planner is to maximize $S_2 = \int_{\theta_2}^{1} U_2(\theta) d\theta$ by choosing A_1. As discussed in Section V.B, after the decision of A_1, the social planner will choose the user adoption such that all users participate in the production with non-negative utility.

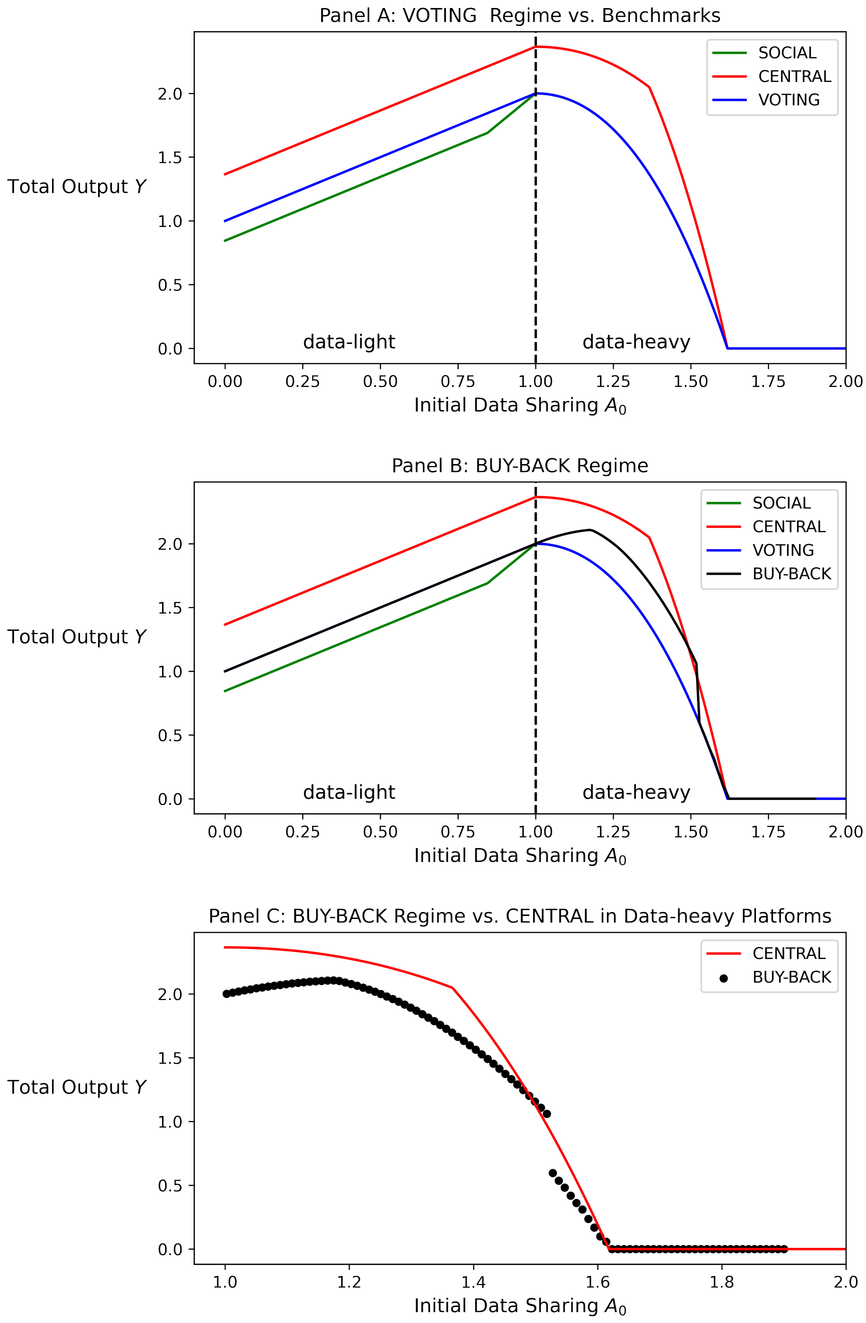

Figure 9. Total output for different governance regimes (parameters: $\overline{A} = 2$, $\beta = 1$)

This figure compares the total output of four distinct governance regimes. Following proposition 5 in Section V.C, Panel A suggests that the voting-only regime sits between the centralized platform and the social planner's problem in terms of total output. Panel B displays the total output for both 'data-light' and 'data-heavy' platforms. Panel C highlights the efficiency of the buy-back and voting regime relative to centralized governance in 'data-heavy' platforms.

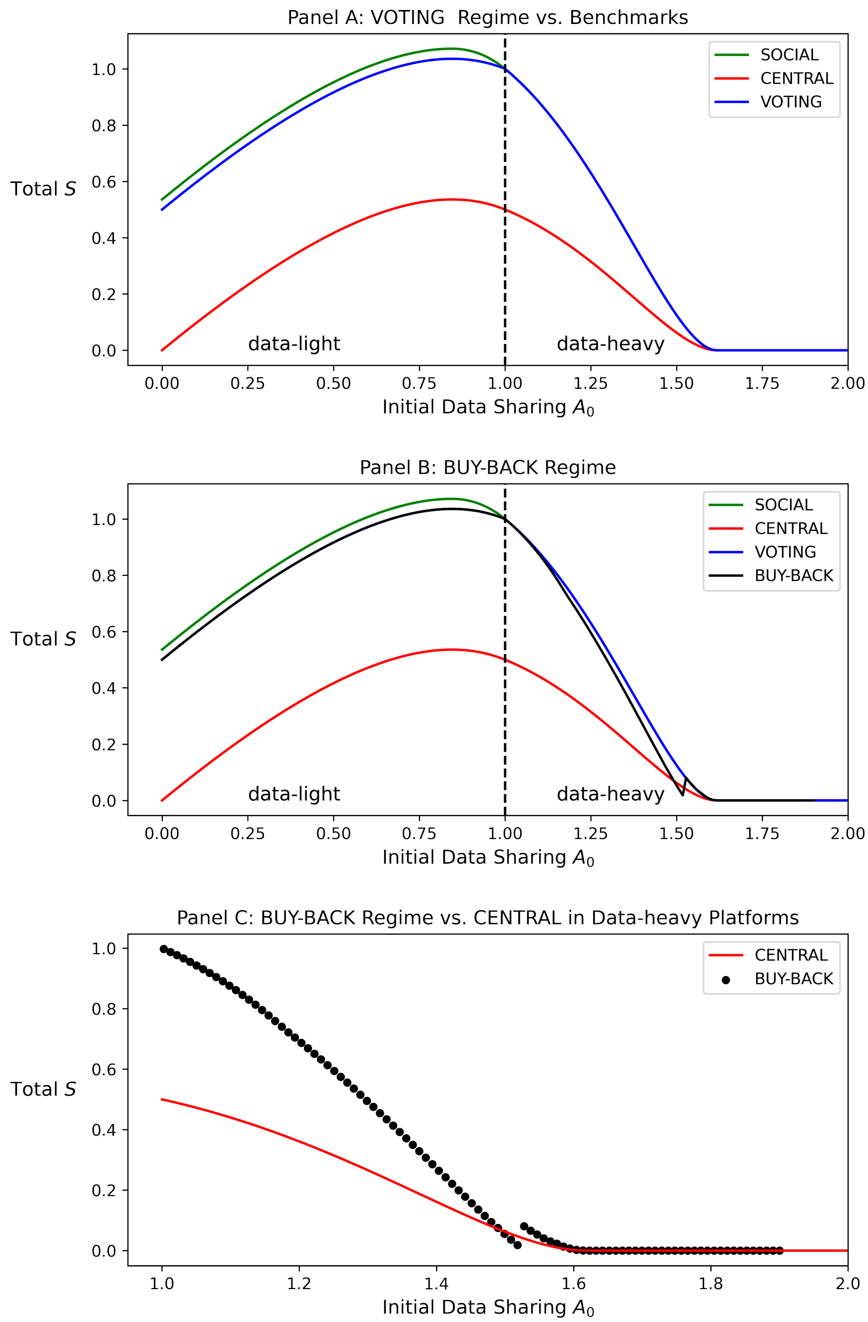

Figure 10. Total user surplus for different governance regimes (parameters: $\overline{A} = 2$, $\beta = 1$)

This figure compares the total user surplus of four different governance regimes. Following proposition 6 in Section V.D, Panel A suggests that the voting-only regime is in the middle of the centralized platform and the social planner's problem. Panel B displays the total user surplus for both 'data-light' and 'data-heavy' platforms and highlights the fairness of the decentralized platform. Panel C presents the user surplus in 'data-heavy' platforms for buy-back and voting regime and centralized governance.