Digital Currency, Data Monopoly and Extractive Adverse Selection in Payments and Lending*

Itai Agur
IMF, iagur@imf.org

Anil Ari
IMF, aari@imf.org

Giovanni Dell'Ariccia
IMF, gdellariccia@imf.org

November 15, 2022

Abstract

The issuance of digital currencies by BigTechs raises the specter of unprecedented data monopolies. This paper examines the difference between the privately and socially optimal extent of data commercialization in a random-state model where a digital currency competes for users who differ in the degree to which they value privacy and have incentives to disclose hidden qualities. Households with better quality and lower privacy preferences opt for increased disclosure, and in doing so worsen the borrowing conditions of other households, due to a Lemons' Market effect. Socially optimal regulation trades off the benefit of limiting intrusiveness against the value of type disclosure in enabling financial access.

Keywords: Privacy; Cascades; BigTech; FinTech; Digital currency regulation.

JEL codes: D82, E41, G21, G28

*This paper is work in progress. We would like to thank Michael Junho Lee (conference discussant), Sole Martinez Peria, Divya Kirti and Germán Villegas Bauer for discussions and audiences at IMF's 5th Annual Macro-Financial Conference and an internal IMF seminar for helpful comments. The views expressed are those of the authors only and do not represent the views of the IMF, its Executive Board or IMF management. Postal address (all authors): IMF, 700 19th St NW, 20431, Washington DC.

1 Introduction

The rise of BigTechs has fueled information monopolies and increasingly brought to a head the tradeoff between privacy and the benefits that information disclosure through new technologies can bestow on those who participate: allowing one's behavior to be watched and parsed by digital eyes can be simultaneously displeasing and rewarding. In few realms is this tradeoff more acute than in payments, because the custodians of payment data have households' economic behavior in a direct line of sight.

A digital currency (DC) is an ideal instrument for a private company to monopolize payment data, because a currency is a walled garden of exchange transactions, especially when the same company that issues the currency also provides the platform for its use. For instance, Meta Platforms has twice proposed (through the Libra and Diem initiatives) the issuance of its own DC, which could have enabled it to turn its enormous network into a payment environment. Although Meta's initiatives did not materialize, the evolutions of BigTechs and digital currencies may not remain separate for long (Brunnermeier et al., 2021). A private company that couples a business model based on the collection and sale of data with the natural monopoly that is currency would attain unprecedented information and market power. It is crucial for policy makers to understand the welfare implications of such a development and whether and to what extent regulation should step in to control it. This paper builds a framework to scope this question.

We distill currency selection as a choice between making payments with cash or a DC that is issued by a private company. Beyond the where and when of a transaction, a DC that is linked to an ecommerce or social media platform can also reveal its what or why. Even so, the issuer of the DC can choose to monetize only certain segments of its vast trove of user information, for instance through the exclusion of social media interactions between payer and payee. We model this with a probability of type revelation: a more intrusive DC design increases the chance that the credit quality (discussed below) of a DC user is revealed.

Our model contains households, lenders and a DC issuer.[1] Households are both consumers and entrepreneurs. As consumers, they differ in the privacy cost they attach to paying with an instrument that observes them intently. As entrepreneurs, they each own a project that needs financing to come to fruition and these projects differ in their probability of success, meaning households are heterogeneous in their credit quality. Projects yield no payoff if they are not financed or if they are financed but fail. The payoff of successful projects depends on the aggregate economic state. In a good state, returns on successful household projects are higher.

The net return on household projects also depends on which lender they borrow from. Each lender is a "home" lender to a subset of households, which can broadly represent industry or geographic specialization. Households can borrow from other lenders too, but their projects are most profitable when matched with the home lender. Lenders obtain their funds at a fully elastic market rate and engage in Bertrand competition on loan rates to households. Here, lenders have two sources of market power: home lender advantage and information. It is insight into household credit quality that enables customized loan rates and such insight is obtained by access to the data that the DC issuer collects.[2]

The DC issuer maximizes its profits by designing the DC (determining its degree of intrusiveness) and deciding on the data access fees to lenders.[3] The DC issuer makes these choices after the aggregate state is revealed, but before households sort into means of payment, lenders decide on data acquisition and loan rates, and households borrow to finance their projects. As part of its optimization, the issuer considers how its DC design affects household choice to sort into DC and cash use: households with relatively low privacy costs

[1] In the baseline model the DC issuer acquires and sells data, but Appendix E [to be done] extends to a setting where the DC issuer also engages in credit provision.

[2] Appendix D considers how the model is affected if lenders have access to other sources of information on credit quality, such as credit registries.

[3] Selling its data through non-exclusive contracts is optimal for the DC issuer here. As shown in Appendix E [to be done], the DC issuer would be worse off if it provided data exclusively to one lender, either by selling to or owning one lender, as foresight of higher loan rates by exclusive lenders would dissuade too many households from using the DC.

and good credit quality have the largest incentives to disclose and thus opt for DC use.[4]

In equilibrium, the DC issuer prices data access to the home lenders just below the value added of all DC users' projects, while granting other lenders free access to its data. Home lenders do not refuse to pay, because adverse selection among revealed borrowers precludes a profitable alternative for uninformed home lenders. The DC issuer combines this strategy with a maximally intrusive DC design, which purposely instigates a Lemons' Market effect. As more DC users' credit qualities become revealed, the expected credit quality of all unrevealed households (namely, cash users and those DC users that remain unrevealed) decreases, and therefore the loan rate on this pool of unrevealed households rises. This induces some households of relatively good credit quality to attempt disclosure through DC use and their type revelation further raises loan rates on the unrevealed, inducing a cascade of disclosure that plays into the DC issuer's hands.

Disclosure cascades endow privacy with a public good nature in this model. Hence, a more intrusive DC design not only raises privacy costs but may also bring about negative externalities by eroding the value of this public good: others' actions to forego privacy make it costlier (in terms of loan rates) for an individual household to opt against disclosure. However, the ability of households to signal their types can also create positive social value to DC use by enabling positive net present value (NPV) projects to receive financing.

The conflicting social welfare implications of DC come to the fore when we investigate equilibria under good and bad aggregate states. In a good state with high returns on successful projects, there are relatively few bad borrowers whose projects are expected to yield less than the financing needed to get them started. In this state, bad borrowers obtain credit in the pool of the unrevealed because they share the pool with enough good borrowers to sustain credit provision. A more intrusive DC design then has no impact on aggregate

[4]Cash here represents a payment technology that completely prevents type disclosure and its comparison to DC provides for a stark setting to highlight the model's key mechanisms. However, the crucial assumption here is that DC use reveals *more* about household credit quality than alternative payment technologies, not that the alternative technologies are entirely unrevealing, as shown in Appendix D.

credit, but reduces welfare by raising privacy costs and externalities.[5] In contrast, in a bad state where returns on successful projects are low, the loan market for unrevealed households freezes.[6] A more intrusive DC now gives households with relatively good credit quality a more potent tool to stand out and obtain credit. This raises aggregate credit and welfare despite increased privacy costs.

We use this framework to draw insights about optimal DC regulation.[7] We assume that regulation is designed at the beginning of time, before the aggregate economic state materializes, which speaks to the notion of DC regulation as a slow-moving policy that is determined with the long view in mind. Optimal regulation trades off privacy costs and externalities against the benefits of disclosure at times when being able to signal quality enables credit provision. Depending on the probability that different states occur and the project returns in those states, optimal regulation can tend to the corners (banning DC or laissez-faire) or an interior solution where regulation restricts the intrusiveness of the DC.

Honing in on the latter, we analyze how the distribution of project returns affects the distance between privately and socially optimal DC design. Increasing the mean of the distribution of project returns affects optimal regulation non-monotonically. As mean returns rise, optimal regulation first becomes looser because higher project returns increase the social value of disclosure in the bad state; however, beyond a threshold, an increase of mean returns leads the regulator to ban the DC because the social value of disclosure in the bad state dissipates, as credit access no longer requires type revelation.[8] The baseline model considers

[5] Households with high credit quality are better off when the DC is more intrusive and type revelation becomes more likely, but this is a distributional effect only: their gains from lower loan rates are exactly offset by the losses from higher loan rates among the unrevealed.

[6] We also investigate intermediate states in which DC design determines whether lending to the unrevealed is sustained: a sufficiently intrusive design collapses that market.

[7] This question is at the forefront of policy debates, with a particular focus on the potential role of BigTechs in payment and credit provision (Adrian, 2021; Bains et al., 2022; Boissay et al., 2021; IMF, 2020).

[8] We also analyze the impact of a mean-preserving spread, whereby returns in good and bad states symmetrically increase and decrease. The impact of this spread depends on the mean. Starting from a mean that is low enough that the unrevealed market is frozen in the bad state, an increase in the spread leads to more stringent DC regulation and a larger gap between privately and socially optimal design. Because lending is available to all in the good state, returns in the bad state dominate the social planner's considerations. The value of information revelation falls when there are fewer good quality borrowers in the bad state.

a binary state realization, but Appendix B extends to a full distribution of states, where optimal regulation no longer exhibits discrete jumps, while the direction of comparative statics is otherwise unchanged [*to be confirmed*].

One way to interpret these results is in terms of cross-country differences. On the one hand, developing economies typically experience greater macroeconomic volatility than advanced economies, and can be represented by a higher variance of returns, which would correspond to tighter optimal DC regulation. On the other hand, in developing economies, larger shares of the population tend to lack financial access and therefore stand to benefit from instruments to build up credit profiles with.[9] In our framework this can be represented by a higher probability of the bad state, which implies looser optimal DC regulation. Hence, our model can be used to highlight the interplay of various factors that determine the optimal extent of DC regulation in different economies.

Our paper relates to a growing literature on emerging financial technologies in payments and credit provision.[10] A common thread in various papers in this literature is that giving households ownership of their data can be socially suboptimal. In Parlour et al. (2022b), this derives from the social value of payment data in credit provision: the entrance of FinTech payment providers can disrupt the information monopoly of a bank that couples insights garnered from payment services with its credit provision.[11] In He et al. (2022), the sub-optimality of household data ownership arises from the fact that individual data sharing decisions lead to aggregate credit quality inferences by lenders. To households, this can outweigh the benefits from intensified loan market competition between banks and FinTechs

[9] See, e.g, Cornelli et al. (2020), Croxson et al. (2022), Frost et al. (2020), Huang et al. (2020), Mester (2020) and Ouyang (2022).

[10] Much of this literature builds on the classical literature on optimal disclosure (Akerlof, 1970; Grossman, 1981; Grossman and Hart, 1980; Jovanovic, 1982; Milgrom, 1981, 2008), and also stands at the intersection of the Kocherlakota (1998) "money is memory" (sharable transaction histories in the context of this literature) versus Kahn et al. (2005) "money is privacy" debate.

[11] On market power, asymmetric information, and data externalities, see also Bergemann et al. (2022), Broecker (1990), Choi et al. (2019), Cicala et al. (2021), Gambacorta et al. (2022), Garratt et al. (2021) and Parlour et al. (2022a).

that is facilitated by data sharing.[12] In Markovich and Yehezkel (2021)'s analysis of a digital platform's data commercialization, it is heterogeneity in the privacy costs of platform users that underlies the suboptimality of giving them ownership of their data, because too many users then underprovide data, which comes at a public cost as it worsens the functioning of the platform for all users.[13,14]

Several papers consider whether the introduction of electronic cash in the form of a Central Bank Digital Currency (CBDC) can be a more effective policy than giving households ownership of their own data.[15] In Garratt and Van Oordt (2021), the social value of electronic cash derives from the fact that privacy is a public good: taking actions to protect digital privacy is costly to a household, which fails to internalize the benefit to others of preventing merchants from learning the relationship between observable characteristics and consumer types. In Garratt and Lee (2021), the introduction of electronic cash raises welfare compared to household ownership of their data: either policy can fend off the endogenous formation of data monopolies, but household data ownership leads to data underprovision, which hampers firms' ability to match products to consumer preferences.[16] In Ahnert et al. (2022) and Brunnermeier and Payne (2022), a CBDC can be more than an electronic equivalent of cash: it can include data sharing features that help achieve the efficient allocation by breaking down private payment data monopolies.

Many of these studies focus on binary policy questions, such as whether to legislate

[12] In Huang (2021) banks also compete with FinTech lenders, which rely on data from linked ecommerce platforms while banks rely on physical collateral, leading to different borrower type specializations for FinTechs and banks. In Fishman et al. (2020) banks choose whether to pay to screen out unprofitable borrowers and this decision imposes dynamic externalities: tighter screening worsens the pool of potential borrowers, increasing banks' incentives to screen in the future.

[13] On ecommerce data collection and privacy, see also Charlson (2021), Chiu and Koeppl (2022) and Ichihashi (2022).

[14] For empirical analyses on the impact of implemented data regulation policies on observed digital privacy outcomes, see Canayaz et al. (2021) and Dorfleitner et al. (2021).

[15] See Agur et al. (2022) for a model where a CBDC's degree of cash-like anonymity is a design choice and see references therein for a review of the broader literature on CBDC. We also note that in our framework, the introduction of a CBDC with a socially optimal degree of data revelation can be an (equivalent) alternative to DC regulation. Croxson et al. (2022) discuss CBDC as a policy alternative to DC regulation.

[16] The potential for suboptimality of data ownership policies is therefore robust to the direction of data externalities (positive in Markovich and Yehezkel (2021) and Garratt and Lee (2021) and negative in most other papers cited here).

consumer data ownership and whether to introduce electronic cash to help safeguard digital payment privacy.[17] This contrasts with countries' approaches to data sharing and privacy regulation, which are typically non-binary. For example, DC regulation is unlikely to tend to outright bans or an entirely laissez-faire approach in most cases (FSB, 2021, 2022; IMF, 2020, 2021, 2022). Our paper contributes to the literature by developing a model that permits interior choices of DC intrusiveness. This yields a rich tradeoff between data disclosure and privacy costs in an environment with financial frictions, and leads to important insights about socially optimal DC regulation and the conditions under which it may constrain privately optimal DC design. A key insight from our model is that socially optimal DC regulation between the extremes is not a simple extension of binary choice and can respond non-monotonically to economic conditions. The reason for this non-monotonicity is that the benefits and costs of tighter DC regulation materialize in different states of the world and therefore optimal regulation responds to the full distribution (including mean and variance) of states.[18]

Our paper also draws from a sizable empirical literature, surveyed in Acquisti et al. (2016) and Uno et al. (2021), which aims to measure preferences and tradeoffs in the realm of digital privacy. Based on survey and experimental data, Bijlsma et al. (2021a,b), Borgonovo et al. (2021) and Hu et al. (2021) find a central role for privacy preferences in payment choice, while Škrinjarić et al. (2022) report extensive heterogeneity in digital privacy preferences among surveyed consumers. Our model incorporates this in a stylized form with a uniform distribution over privacy costs. This also aligns with another finding in the empirical literature that additional dimensions of agent heterogeneity (in addition to hidden quality) avoid a full "unraveling" equilibrium where all agents opt to disclose private information (Jin et al.,

[17] See Piolatto and Schuett (2022) for the discussion and analysis of another digital data policy: the prohibition of price-parity clauses, whereby ecommerce platforms tie sellers' hands to not offer their products at lower prices outside the platform.

[18] There are negative externalities associated with a household's type revelation only when the unrevealed pool receives loans. If the unrevealed loan market is shut, the ability of households to differentiate instead enables financial access.

2021; Jin and Vasserman, 2021).[19]

The remainder of the paper is organized as follows: Section 2 presents the setup of the model, including agents' actions, payoffs, and the timing of the game between them. Section 3 derives the subgame-perfect equilibria and discusses agents' optimal strategies in these equilibria. Section 4 analyzes welfare and socially optimal policy design. Section 5 concludes. Proofs and extensions can be found in the appendices.

2 Model

Our model consists of three forms of agents - households, lenders and a digital currency (DC) issuer - which are all risk neutral. Households may use two payment technologies in their transactions: a 'digital currency' (DC) that has the potential to reveal private information about its users and another payment technology that reveals no information other than its use (which for simplicity we refer to as 'cash').[20] Sections 2.1-2.2 describe the agents, whose actions are summarized in Figure 1. Section 2.5 lays out the timing of the game between them, which Section 3 subsequently solves by backward induction.

2.1 Households

Each household is born with an investment project that can only be brought to fruition with financing from a lender. There are $N \geq 3$ lenders and each household has an assigned "home" lender. Borrowing from the home lender comes with relative benefits compared to borrowing from other lenders, which can be taken to represent, for example, spatial differentiation or industrial specialization of lenders that are matched to household projects in the same industry. The notion of a home lender facilitates the introduction of a degree of

[19] Furthermore, there is empirical evidence that credit quality information garnered through FinTech based payments plays a signficant role in credit provision (Agarwal and Assenova, 2022; Babina et al., 2022; Beck et al., 2022; Ghosh et al., 2021), as it does in our model. On the relation between other (non-payment) data collected by FinTechs and credit market outcomes, see Di Maggio et al. (2021).

[20] The model generalizes to partial revelation from other sources than DC use, as discussed in Appendix D, as long as DC use entails additional potential revelation.

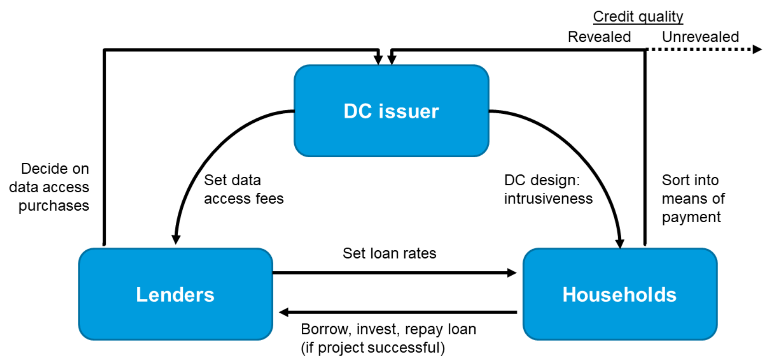

Figure 1: Overview of agents and actions.

lender market power in a tractable form. There is a unit continuum of households assigned to each lender (as home lender) and therefore there is a total mass of $N \times 1$ of households. Which households are assigned to which lender as their home lender is public information.

Successful projects yield an identical payoff y while unsuccessful projects yield zero payoff. The probability that a project succeeds, q, differs across households but is independent of their project's type. We refer to q as a household's credit quality. For each of the N home lenders, there is a mass 1 of households that is distributed uniformly on $q \in \left[\frac{1}{2}, 1\right]$.[21]

Credit quality is private information but it can be revealed through households' use of payment technologies. In particular, in addition to the project, households are born with an endowment that they use towards consumption. Households choose which payment instrument to use for consumption: a digital currency, which reveals a household's credit quality to the DC issuer with endogenously determined probability $\theta \in [0, 1]$ (see Section 2.2), and cash, which is unrevealing and thus has $\theta = 0$.

Revelation imposes a privacy cost that is heterogenous across households. In particular, each household attaches a disutility parameter φ to revelation with households distributed

[21] We choose $q \in \left[\frac{1}{2}, 1\right]$ instead of $q \in [0, 1]$, as the latter creates various additional complexities (individual breakeven loan rates go to infinity as $q \to 0$, which complicates the identification of simple parameter conditions to characterize equilibria (shown in Table 1); for technical detail, see, e.g., footnotes 59 and 60). Using $q \in \left[\frac{1}{2}, 1\right]$ is without loss of generality, however, in the sense that all possibilities on the functioning of the loan market are covered by our framework, as will be seen from the propositions in Section 3.

uniformly on $\varphi \in [0,2]$.[22] These privacy preferences are private information and independent of other household characteristics (i.e., q and φ are independent).[23] Hence, for each of the N home lenders, there is a unit mass of households, which is uniformly distributed on the $\left(q \in \left[\frac{1}{2}, 1\right], \varphi \in [0,2]\right)$ plane.

While households may borrow from any lender type, borrowing from a lender which is not their home lender reduces the payoff of a successful project by $\tau > 0$.[24] The expected payoff for of a household of type (q, φ), considering both the potential gains from borrowing and the costs of privacy, can then be written as

$$u(q, \varphi) = q \max\{(y - \tau I - R), 0\} - \alpha\varphi\theta \tag{1}$$

where the term $q \max\{(y - \tau I - R), 0\}$ represents the expected benefit from borrowing and investing and $-\alpha\varphi\theta$ is privacy costs.[25] In $q \max\{(y - \tau I - R), 0\}$, I is an indicator variable that is equal to 1 if a household borrows from a lender that is not its home lender and 0 when the household's project is matched to the home lender; R is the gross loan rate charged to a household; and the max operator denotes that a household will only choose to engage in borrowing and investing when it can obtain a loan at a sufficiently low interest rate.[26] In $-\alpha\varphi\theta$, $\alpha > 0$ is the relative weight on privacy preferences and $\theta \in [0,1]$ represents the

[22] To be precise, for each of the N project types, there is a mass 1 of households that is distributed uniformly on $\varphi \in [0,2]$. We let $\varphi \in [0,2]$ instead of $\varphi \in [0,1]$ to obtain a unit mass of households for each project type.

[23] [To be done]: We may consider an extension where q and φ are correlated. To retain tractability, a perfect positive ($q = \frac{\varphi+2}{4}$) or negative ($q = \frac{4-\varphi}{4}$) correlation could be considered.

[24] To guarantee that limit pricing can be sustained as a Nash equilibrium in Section 3, we additionally assume a breakeven preference in favor of the home lender. I.e., when loan rates are such that households are indifferent between borrowing from the home lender or other lenders, they choose to borrow from the home lender. An alternative assumption is the approach of Blume (2003), which allows for limit pricing to be supported as a mixed-strategy Nash equilibrium where the home lender charges the limit interest rate and the others mix over a narrow interval around the breakeven rate. But we focus on pure strategy solutions here and therefore work with the breakeven preference.

[25] In the baseline model, privacy costs are therefore linear in the probability of revelation, θ. Appendix C considers quadratic privacy costs.

[26] If the available loan rates exceed the payoff when the project succeeds, the household does not invest and therefore the term $q \max\{(y - \tau I - R), 0\}$ collapses to 0. Note also that households only care about the payoff when the project succeeds, as they have limited liability and do not repay their loan when the project is unsuccessful.

probability that the household's credit quality will become revealed.

Overall, household choice in our framework centers on both payments and borrowing, as households choose which payment instrument to use for consumption and whether and from which lender to borrow. We shall show below that the parameter θ, which represents the intrusiveness of DC (and hence the revelation probability associated with it) will affect both of these decisions.

2.2 Digital currency issuer

The DC issuer collects credit quality data on DC-using households and puts these data up for sale to lenders.[27] The DC issuer's choice variables are therefore $\theta \in [0, 1]$ (which, as explained above, determines the intrusiveness of the DC and the revelation probability associated with it) and the data access fees charged to lenders. We assume that the DC issuer may charge differentiated fees to lender, and that sales are non-exclusive, meaning that the DC issuer may sell data on a given group of households to any lender that is willing to pay for access.[28]

We let Ω_{ij} denote the data access fee that the digital currency provider charges lender i for access to data on households with home lender j. The numbering of lenders by i and household home lenders by j is symmetric. For example, Ω_{11} is the data access fee for lender 1 with respect to households that have lender 1 as its home lender; Ω_{1j} for $j > 1$ are the fees for lender 1 to access the credit quality data of households that do not have it as their

[27] The DC issuer markets the credit quality data of revealed households. An underlying assumption is that it cannot additionally market a list of which households open a DC account. Allowing for that option would lead to two separate pools of unrevealed households: cash users and DC users that remain unrevealed with probability $1 - \theta$. This would open the model to household mimicking strategies, considerably complicating the analysis (i.e., some low q, high φ households may gamble on DC use in the hope that they end up among the $1 - \theta$ unrevealed DC users, implying no privacy cost and lower equilibrium loan rates than among the pool of cash users). However, a minor addition to the model would restore our baseline setup: allowing households to open *unused* DC accounts (while consuming with cash). Opening a DC account and not using it would directly sort a household into the pool of unrevealed DC users. With this option, all cash users would open an additional unused DC account, as it would come without privacy cost but with the advantage to be pooled with unrevealed DC users that have a higher average credit quality and therefore lower loan rates (see the discussions on equilibrium household sorting in Section 3). There would then be the same single pool of unrevealed households as in our baseline model and all results would be the same.

[28] Appendix E [to be done] shows that this the optimal way for the DC issuer to market its data in our framework. Offering exclusive data contracts (i.e., for the highest bidding lender) or owning a lender and providing it with exclusive data access, reduces the DC issuer's profits.

home lender.[29]

We next define $D_{ij}(\theta, \Omega_{ij})$ as lender i's demand function for data on households with home lender j. This demand function will take value 0 (do not purchase) or 1 (purchase) depending on the pricing of the data access, Ω_{ij}, and the value that a lender derives from the data, which will depend on the DC's intrusiveness, θ. That is, how much a lender is willing to spend on data gathered from DC users, will depend on how much borrower data the DC issuer obtained, which in turn relates to the intrusiveness of the DC.[30] Therefore, D_{ij} is written as a function of (θ, Ω_{ij}).

The DC issuer's objective is to maximize data fee revenue by optimally choosing θ and Ω_{ij}.[31] This can be written as

$$\max_{\theta \in [0,1], \Omega_{ij}} \sum_{i \in [1,N]} \sum_{j \in [1,N]} \Omega_{ij} D_{ij}(\theta, \Omega_{ij}) \tag{2}$$

2.3 Lenders

Lenders in our model have an elastic supply of funding at cost $c \geq 1$ and engage in Bertrand competition. They have two sources of differentiation. First, $\tau > 0$ parameterizes the market power derived from home lender advantage. A second potential source of differentiation among lenders is their access to information about households' credit quality. If one lender purchases detailed data on a set of households, while other lenders do not, the purchasing lender can charge differentiated loan rates to attract the subset of households with high credit quality (i.e., a high probability of success q), leaving other lenders a pool of lower credit quality borrowers.

Lender i's profit per borrower is given by two terms. The first concerns the expected profits that the lender makes on revealed borrowers, k, whose credit quality data the lender

[29] The DC issuer therefore sells the data by block of households. That is, for each block of households with specialization j, it markets the data non-exclusively to the lenders. Appendix E [to be done] shows this is optimal for the DC issuer.

[30] This becomes clear from the timing of the game laid out in Section 2.5.

[31] We abstract from costs associated with setting up or managing a digital currency here, because these would not be material to the analysis of a monopolistic DC issuer.

has obtained, and who have chosen to borrow from lender i. We denote such households as ki borrowers. Here, ki can be represented as a set with continuous support on q_{ki}.[32] A lender's expected profit on a loan to a ki borrower, with customized loan rate $R_{ki}(q_{ki})$, is $q_{ki}R_{ki}(q_{ki}) - c$. The second term concerns the expected profit that the lender makes from unrevealed borrowers. For every loan that the lender provides to an unrevealed borrower, the expected profit is $E[q|u]R_{ui} - c$ where R_{ui} is the single loan rate that lender i charges unrevealed borrowers, and $E[q|u]$ represents the expected credit quality of a household conditional on being unrevealed. Here, ui denotes the unrevealed households that choose to borrow from lender i. Aside from these expected profits per borrower, the lender also considers data acquisition costs if it chooses to purchase data on revealed borrowers from the DC issuer, as per (2).

Overall, lender i's expected profits can be written in general form as

$$m_i\left(E[q_{ki}]R_{ki}(q_{ki}) - c\right) + \upsilon_i\left(E[q|u]R_{ui} - c\right) - \sum_{j\in[1,N]}\Omega_{ij}D_{ij}(\theta,\Omega_{ij}) \tag{3}$$

where m_i and υ_i represent the masses of, respectively, revealed and unrevealed households who choose to borrow from lender i. These, as well as the expectation and loan rate terms, obtain closed forms as part of the derivations of Section 3.

In sum, lenders' decisions center on whether to purchase detailed data on households and what loan rates (or loan rate schedules where credit quality data is purchased) to charge. In the data purchase decision, lenders are modeled as price takers: they are made offers by the DC issuer which they can take or leave.

2.4 Parameter conditions

We include three parameter conditions, which relate y, τ and α to c, the lenders' cost of funding that is further discussed in Section 2.3. First, we note that lenders may encounter

[32] The Proof of Proposition 1 shows that this is an accurate representation when solving the game in Section 2.5 by backward induction.

both positive and negative NPV borrowers in our framework. In particular, we let

$$y > c \tag{4}$$

which ensures that the highest quality borrower ($q = 1$) always has a positive NPV project.[33]

Second, we let

$$\tau < \frac{1}{3}c \tag{5}$$

which means that the lost revenue on successful projects that households suffer by not borrowing from their home lender is at most $\frac{1}{3}$ of lenders' funding cost. Key mechanisms of our model rely on the competition between lenders and when home lender market power becomes too strong, such mechanisms break down.[34]

Third, we assume that

$$\alpha < \frac{1}{6}c \tag{6}$$

which helps preserve tractability.[35,36]

2.5 Timing

The timing of agents' actions and the realization of events is as follows:

1. The DC provider determines DC design, θ, and data access fees, Ω_{ij}.

2. Households sort into DC/cash users and consume.

3. DC users become revealed with probability θ.

[33] This is a necessary condition for credit provision to take place.

[34] See the Proof of Proposition 1.

[35] Technically, this condition is necessary and sufficient to ensure that households with the highest credit quality ($q = 1$) will always want to disclose their type, regardless of their privacy preferences. See the discussion in the paragraph of equation (20) for details. Relaxing (6) has been investigated, but leads to highly complex expressions that do not readily lend themselves to analysis.

[36] A sufficient condition for (6) is $\alpha < \frac{1}{6}$ since $c > 1$. While our model is not quantitative in nature, it seems likely that α, which parameterizes privacy disutility versus consumption preferences, will in practice be relatively small (Athey et al., 2017).

4. Each lender (i) decides for each set of households, with home lender j, whether to purchase data access at fee Ω_{ij}.

5. Each lender announces its loan rates (customized loan rates for revealed households on which data has been purchased and a single loan rate for other households).

6. Households decide whether to borrow and, if borrowing, which lender to borrow from. Households that borrow invest in their projects.

7. Project returns are realized and households that invested in successful projects repay their lenders.

In this game, the DC issuer is the first mover: the way it designs its DC feeds into the rest of the game. In particular, based on the DC design, households decide whether to use DC or cash for their consumption. After some households' credit qualities become revealed to the DC issuer, lenders must decide whether to purchase data. Lenders then determine individualized loan rate schedules for households for which they know the credit qualities, and offer a single loan rate for other households. Finally, households decide whether to invest in their projects and if so, from which lender to borrow. Households with successful projects then repay their lenders in the last stage. In the next section, we work backward through the stages of the game to derive the equilibria, based on which we further elaborate on the optimal strategies of households, lenders and the DC issuer.

3 Equilibria

When the payoff on a successful project y is only barely above the lender's cost of funding c, then households will have negative NPV projects ($qy < c$) unless their probability of success q is close to 1, leading to a large share of "bad borrowers" on the loan market. Conversely, when y is well above c, households will have positive NPV projects ($qy > c$) unless their probability of success q is very low, leading to a large share of "good borrowers" in the loan market. From hereon in, we will refer to borrowers with positive (negative) NPV projects as good (bad) borrowers.

Below we derive the subgame-perfect equilibrium of the game laid out in Section 2.5 in three separate propositions.[37] These propositions are defined for different cases, as delineated by the parameter y. Which types of lending markets are tenable under asymmetric information relates to the share of bad borrowers and therefore equilibrium outcomes can differ depending on y. We define the following three cases:

Table 1: Parameterizing the share of bad borrowers

Case	Share of bad (NPV < 0) borrowers	Value of y
High y	Small	$y > 2c + \tau$
Medium y	Intermediate	$y \in \left(\frac{4}{3}c, 2c + \tau\right)$
Low y	Large	$y \in \left(c, \frac{4}{3}c\right)$

3.1 Equilibrium with few bad borrowers

In the case with high y where the share of bad borrowers is small, the equilibrium is described by the following proposition:

Proposition 1 *When $y > 2c + \tau$: there is a unique subgame-perfect Nash equilibrium in which: at Stage 1, the DC issuer sets $\theta = 1$ and $\Omega_{ij} = 0$ when $j \neq i$ and $\Omega_{ij} > 0$ (with the closed form solution for the optimal fee given in (33)) when $j = i$ (i.e., the fee for lenders to access the data on households to which they are home lenders); at Stage 2, households with*

$$\varphi \leq \frac{c}{\alpha}\left(\frac{q}{E\left[q|\, u\right]} - 1\right) \tag{7}$$

sort into DC use and households above this threshold sort into cash. Here the closed-form solution for $E\left[q|\, u\right]$ is displayed in (30) and, replacing this, (7) is in closed-form too. At

[37] In our derivations below, we concentrate on symmetric pure-strategy equilibria among the lenders. However, in footnotes within the proofs of the respective propositions, we show that the results extend to mixed-strategy lender actions if N is large enough.

Stage 4, each lender buys the data on households that have it as a home lender and at Stage 5 such lenders offer differentiated loan rates

$$R_k(q) = \frac{c}{q} + \tau \tag{8}$$

to revealed borrowers and a single loan rate

$$R_u = \frac{c}{E[q|\, u]} + \tau \tag{9}$$

to unrevealed borrowers.[38] *There is a closed-form solution for R_u, displayed in (34), and its comparative statics to the underlying parameters are shown in (35); at Stage 6, all households borrow from their respective home lenders and invest in their projects.*

Proof. See Appendix A (p.37). ■

Households' optimal strategies

An important aspect of condition (7) is that R_u depends on the credit quality of all unrevealed households, as can be seen in (9). Therefore, the payment choice of a given household is affected by the payment choice of other households. In particular, when a larger share of relatively good credit quality households choose to use DC and some of these become revealed, this lowers the average credit quality of the pool of unrevealed borrowers. That, in turn, raises the loan rate that all unrevealed borrowers face. And since R_u is part of (7), a change in R_u feeds back into payment choice: there are feedback effects between the payment choices of households.

Such feedback effects in our model take the form of negative privacy externalities and cascades of revelation. This can be highlighted with a comparison to a cash-only world. If there were no DC, all households would use cash and $E[q|\, u] = E[q] = \frac{3}{4}$, so that $R_u = \frac{4}{3}c + \tau$. Then, imagine the DC is introduced: at $R_u = \frac{4}{3}c + \tau$, some households will find it attractive

[38] We drop the subscripts i in (8) and (9), given home lender symmetry (each lender is a home lender to a given set of households), as further discussed in the Proof of Proposition 1.

to use DC instead of cash. This set of households is represented by the households to right of and below the dashed line in Figure 2. Since these households are located in the corner with the highest credit quality, their choice to use DC (meaning some of them will become revealed) lowers $E\left[q|\, u\right]$ and raises R_u. Then, from (7), a lower $E\left[q|\, u\right]$ implies that more households find that $\varphi \leq \frac{c}{\alpha}\left(\frac{q}{E[q|u]} - 1\right)$ and they choose DC over cash. But these are households that were close to the threshold given by the dashed line, meaning that their credit quality was high among cash users: their choice to use the DC in turn further lowers lowers $E\left[q|\, u\right]$ and raises R_u, and so on. The equilibrium relations in Proposition 1 document the ultimate settle point of such a process, as represented by the unbroken line in Figure 2. But behind this equilibrium lies a cascade of revelation: households who only choose to use the DC and thereby may reveal their credit quality, because other, higher credit quality households choose to do so.

This is a combination of the dynamics of a Lemons' Market and a setting in which privacy is a public good. It is this public good nature of privacy that implies that households can impose negative privacy externalities on each other. Households in the bottom right corner of Figure 2 are surely better off in a world with the DC than if only cash were available. They are of such high credit quality (and may also see limited costs to privacy) that they wish to reveal their types to get lower loan rates. For households in between the dashed and unbroken lines, instead, the story is different. They are swept along by the force of others' choice to reveal and its implications on their loan rates if they do not go along and can therefore be worse off than if the DC did not exist. The implications for individual and aggregate welfare are further explored in Section 4.

Households at the left end of Figure 2 are bad borrowers. Bad borrowers self-select into cash use, because they know that credit quality revelation will preclude them from receiving loans. However, when $y > 2c + \tau$, there are always enough good borrowers on the market, including good borrowers with high privacy preferences that choose to use cash, so that the loan market for unrevealed borrowers remains up and running.

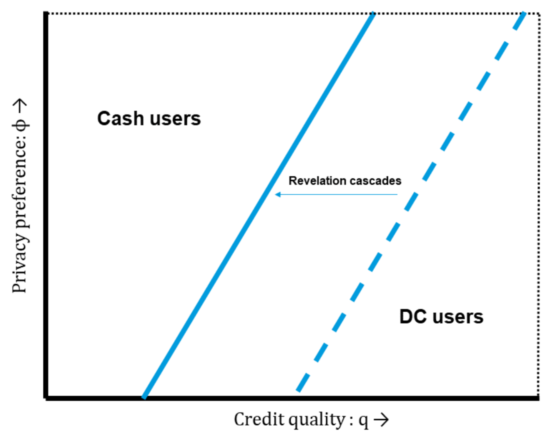

Figure 2: Payment choice and revelation cascades.

Lenders' optimal strategies

Lenders too find themselves swept along by the forces unleashed by information revelation. When all lenders have the same access to information, home lenders have a degree of market power, τ. Lenders engage in Bertrand competition on loan rates, but because households earn τ more on projects financed by home lenders, households are willing to pay up to τ higher loan rates on loans from such lenders and therefore equations (8) and (9) emerge from limit pricing. But these equations are the outcome of not only the pricing game between the lenders, but also of the information acquisition game among them. And in this game, it is the DC issuer that holds the cards and sets the stage in such a way as to maximally extract lender profits.

The DC issuer's optimal strategy

The DC issuer wants the home lenders to provide the loans to their respective households, but only after paying a fee that transfers lender profits on revealed households to the greatest extent possible. The DC issuer knows that loan rates and therefore lender profits per loan are at their highest when households are matched to their home lenders. Moreover, the fees that the DC issuer can maximally charge depend on the value of data on revealed households to the lenders. The DC issuer therefore designs its fee structure in such a way that households do end up borrowing from the home lenders and those home lenders have no choice but to transfer the full profit that they make on revealed households to the DC issuer.

The way the DC issuer achieves this is by offering all the other lenders free access to the data on households of a given home lender. By doing so, it plays on adverse selection effects. If the home lender fails to buy the data on revealed households, then it must offer a single loan rate below (9) in order to attract some revealed borrowers.[39] But when an uninformed home lender cuts its loan rate below (9), it is the lower quality end among the revealed households that will select to borrow from it. The uninformed home lender makes less profits on unrevealed borrowers by cutting its loan rate, while facing adverse selection among the revealed households. Rate cuts end up reducing the home lender's profits and therefore it will choose to instead purchase the data from the DC issuer, as long as it makes any positive profit on revealed households. The DC issuer charges the home lenders a fee marginally below τ times the mass of revealed households times their expected quality ($\tau m E\left[q|\, dc\right]$, solved in closed form in (33)), which equals the profits they make on lending to the revealed households, and the home lenders choose to pay this fee.

To maximize this fee, the DC issuer sets $\theta = 1$, which ensures that all DC users become revealed. Setting $\theta = 1$ also maximizes the mass of DC users. This derives from the fact that the pool of the unrevealed consists of two types of households - cash users and unrevealed DC users - and the worst quality unrevealed DC user is always of better credit quality than the average cash user (as shown in the Proof of Proposition 1).[40] When θ increases, more DC users become revealed, the mass of unrevealed DC users shrinks, meaning the average quality of the unrevealed pool declines. Thus, $E\left[q|\, u\right]$ declines and R_u increases, when θ

[39] If the home lender offered the loan rate in (9), no revealed household would borrower from it, because all DC users have a q that is higher than $E\left[q|\, u\right]$ and are therefore better off borrowing at loan rate $\frac{c}{q}$ from other lenders than at loan rate $\frac{c}{E[q|u]} + \tau$ from the home lender, as shown in the Proof of Proposition 1.

[40] The expected quality of DC users, $E\left[q|\, dc\right]$, does decline when θ increases, which, ceteris paribus, lowers data access fees. However, this effect is dominated by the increase in the mass of revealed borrowers: overall data access fees rise as θ increases.

increases and this induces more households to opt for DC use.[41]

3.2 Equilibrium with many bad borrowers

Proposition 2 *When $y \in \left(c, \frac{4}{3}c\right)$, then in the case where $\tau > y - c$, no credit provision takes place and all households use cash. Instead, as long as τ is small enough ($\tau < y - c$), some high credit quality households (namely those for which $q \geq \frac{\alpha\varphi + c}{y - \tau}$) choose to use the DC and, if revealed, borrow from their home lender with loan rates given by (8). All other households choose to use cash. The DC issuer optimally sets $\theta = 1$ and a positive data access fee to the home lender, the closed-form expression for which is shown in (43), while offering the data for free to the other lenders.*

Proof. See Appendix A (p.46). ■

Households' optimal strategies

When $y \in \left(c, \frac{4}{3}c\right)$, unrevealed households can never obtain loans. Bad borrowers outweigh good borrowers among the unrevealed ($E\left[q|\, u\right]$ is high) and therefore the breakeven loan rate on the unrevealed, $\frac{c}{E[q|u]}$, is too high to induce unrevealed households to participate in the loan market. Hence, in the case with many bad borrowers, there are no revelation cascades. The cascades in Section 3.1 are the result of feedback effects that run through the loan market for unrevealed borrowers. When that loan market is closed, the interaction between the payment choice of one household and other households ceases. The choice of a set of households to use the DC still affects $E\left[q|\, u\right]$, but $E\left[q|\, u\right]$ does not affect the return on choosing to be a cash user, when the unrevealed do not obtain loans. Thus, in the case with many bad borrowers, there are no negative privacy externalities and a household's payment choice is a purely individual decision, unaffected by others. Loan pricing in line with (8)

[41] Appendix C considers how the comparative statics of the model are affected when privacy costs are quadratic instead of linear. For some parameterizations, the mass of DC users then declines as θ increases: the incentive to select cash rises as θ increases, because higher marginal privacy costs from DC use dominate higher loan rates from cash use.

would leave some households enough net return on successful projects that they choose to use DC and, if revealed, borrow from the home lenders.

Lenders' optimal strategies

When $\tau < y - c$, competition among lenders establishes limit pricing in line with equation (8). Hence, high credit quality households can obtain loan offers that are enough to induce a choice for DC over cash.

Instead, if $\tau > y - c$ then home lenders effectively become monopolists versus revealed households: such households may have positive NPV projects when borrowing from the home lender, but always have negative NPV projects when borrowing from other lenders. However, due a time inconsistency problem, the effective monopoly of home lenders actually leads to a credit market freeze in this setting. At Stage 5 of the game in 2.5, the home lender will charge revealed households customized loan rates $R_k(q) = y$ that transfer all the profit of a successful project to the lender. Anticipating this at Stage 2, no household will choose to use the DC, because there is a privacy cost to using the DC, but no benefit: the DC enables credit access that however yields a zero return. Lenders cannot precommit to charge lower loan rates and therefore no credit provision takes place when $\tau > y - c$.

As the key insights of this paper center on the interaction between DC use and credit provision, we center attention on the case where home lender benefits are small enough to sustain competition among lenders and assume for the remainder of the paper that

$$\tau < y - c \tag{10}$$

where we note that (10) is tighter than (5) if and only if $y < \frac{4}{3}c$.[42]

The DC issuer's optimal strategy

[42] I.e., the inclusion of (10) only matters in the low y case considered in Proposition 2. Put differently, taken together, (5) and (10) can be written as $\tau < \min\left\{\frac{1}{3}c, y - c\right\}$ or equivalently $\tau < \min\left\{y, \frac{4}{3}c\right\} - c$ where $\min\left\{y, \frac{4}{3}c\right\}$ highlights the relation to the y cases in Table 1.

Given (10), credit provision takes place, and the DC issuer offers household data for free to other lenders. It does so to maximize the pressure on home lenders to pay the data access fee that appropriates all profits on revealed households. Moreover, the DC issuer optimally sets $\theta = 1$, which brings about the largest number of revealed borrowers and thereby maximizes its access fees.

The optimal data access fees charged by the DC issuer ensure that no profit remains on revealed households. In the low y case considered in Proposition 2, adverse selection problems are at their worst: if the home lender refuses to buy the data on revealed households and attempts to charge a single loan rate that attracts some revealed households, all the unrevealed households that are otherwise credit excluded, will rush to borrow from the home lender too. The home lender cannot distinguish the household types and there are too many bad borrowers among the unrevealed, so that the home lender is certain to make a loss if it does not purchase data access.

3.3 Equilibrium with an intermediate share of bad borrowers

Proposition 3 *When $y \in \left(\frac{4}{3}c, 2c + \tau\right)$, then, depending on parameter values, either all households can be serviced in the loan market with outcomes as in Proposition 1 or the loan market for the unrevealed can shut down and only better quality households may be serviced as in Proposition 2. One of the parameters determining which outcome prevails is θ: depending on other parameters, there can be a threshold value, $\widetilde{\theta}$, such that for $\theta < \widetilde{\theta}$ Proposition 1 prevails and for $\theta > \widetilde{\theta}$ Proposition 2. In such cases, Proposition 2 comes about, since the DC issuer continues to optimally set $\theta = 1$.*

Proof. See Appendix A (p.48). ∎

From Sections 3.1 and 3.2, the interaction between y and model outcomes centers on whether the loan market for unrevealed borrowers functions. For $y > 2c+\tau$, there are always enough good borrowers with high privacy preferences among the cash users to ensure that the unrevealed obtain loans. For $y < \frac{4}{3}c$, the opposite is true. Instead, for $y \in \left(\frac{4}{3}c, 2c + \tau\right)$, either can be true, depending on parameters, including θ. For instance, if $\theta = 0$ and all households are unrevealed, then the loan market for the unrevealed functions for any $y > \frac{4}{3}c$. Per $\frac{\partial E[q|u]}{\partial \theta} < 0$ from (30), the higher is θ, the lower is $E[q|\,u]$ and a lower average quality of the unrevealed means that the loan market for the unrevealed comes closer to a freeze, as there are too few good unrevealed borrowers left to cross-subsidize the bad ones. This underlies the notion of a threshold $\widetilde{\theta}$, where $\theta > \widetilde{\theta}$ leads to an equilibrium as in Section 3.2 and $\theta < \widetilde{\theta}$ leads to an equilibrium as in Section 3.1. Since at $\theta = 0$, the unrevealed market always functions here, $\widetilde{\theta} \in (0, 1)$ exists if at $\theta = 1$ the unrevealed do not obtain credit.[43] When that is the case, the equilibrium is as in Proposition 2 because $\theta = 1$ remains optimal for the DC issuer. If the unrevealed loan market functions for any θ (and there is no $\widetilde{\theta}$), then the outcome is given by Proposition 1.

4 Policy analysis

We now draw insights from our model for the analysis and design of socially optimal policy. The type of policy we focus on, is regulation that can constrain the DC's intrusiveness (i.e., the extent of data collection that it facilitates). We denote socially optimal digital currency design by θ^*. In the preceding section, we found that $\theta = 1$ is privately optimal for a DC issuer. Hence, when $\theta^* < 1$, socially optimal policy diverges from privately optimal policy and there is a justification for regulation. In particular, if $\theta^* \in (0, 1)$, it is socially optimal to constrain the extent of digital currency intrusiveness.[44] Instead, if $\theta^* = 0$, it is best to ban the use of DC altogether. The other extreme, $\theta^* = 1$, implies that a laissez-faire policy is socially optimal even in the face of a maximally intrusive DC design.

In order to enable welfare analysis, we extend our model by adding two more stages to

[43] A notable feature of this setting is that as θ increases above $\widetilde{\theta}$, the mass of DC users may either increase or decrease. This is formally shown in footnote 74 and, as discussed there brings together two counterveiling forces in the model: when the unrevealed market shuts, cascades unwind, lowering DC use; but the market shutting also make DC based revealation the only way to obtain credit, raising DC use.

[44] Alternatively, the policy maker could forbid private digital currencies and instead introduce its own DC, such as in the form of a CBDC with design θ^*.

the game described in Section 2.5. These two new stages precede the existing seven stages: First, we assume that the policy maker determines the socially optimal DC design θ^* at the beginning of the game. This is equivalent to assuming that regulatory policies on DC would be infrequently revised, which could emanate from a variety of reasons, including practical difficulties in changing data access and commercialization rights ex-post (i.e., after usage data have already been collected and sold) and the significant up-front costs associated with widespread DC adoption which would necessitate a degree of regulatory certainty.

Second, we let y represent an uncertain state variable: its realization materializes in a second stage, after θ^* is determined. With probability $\gamma \in [0, 1]$, y takes the value y_h and a good economic state materializes. In the good state, credit risks to lenders are limited as there are relatively many good borrowers. This is the case described by Proposition 1. With probability $(1 - \gamma)$, a bad state, y_l, occurs, in which credit risks are profound and there are relatively many bad borrowers. This is the case described by Proposition 2. That is, we let y_h represent "High y" in Table 1, with $y_h > 2c + \tau$, while y_l represent "Low y" in Table 1 with $y_l \in \left(c, \frac{4}{3}c\right)$. In Appendix B [*to be done*], we generalize further to a setup where y is a continuous random variable with support over all three cases in Table 1.[45] At its decision stage, the DC issuer now faces a constrained design choice such that $\theta \in [0, \theta^*]$ and optimally selects as intrusive a DC design as regulation permits, that is, $\theta = \theta^*$.

4.1 Welfare analysis

We define aggregate social welfare as the sum of the expected payoffs (given by equation (1)) of all households plus the total profits of lenders and the DC issuer. Aggregate welfare is then determined by two factors only. The first factor is the value added of the projects that receive funding.[46] The second factor is the privacy costs experienced by DC users. In

[45] We focus on here on the binary (y_h, y_l) case because it is tractable and lends itself to analytical solutions, whereas the general case in Appendix B [will be] numerically solved. Moreover, the spectrum of socially optimal policy outcomes is as wide in the binary as in the general case: there is no apparent loss of generality from focusing on the binary case.

[46] Some of this value added accrues to households, through positive net returns on projects, and some of the value added is captured as corporate profits.

general form, welfare can now be expressed as

$$W = \int_{\text{borrowers}} (qy - c)\, f\,(q|\,\text{"borrowing"})\, dq - \int_{\text{DC users}} (\alpha\varphi\theta)\, g\,(\varphi|\,\text{"DC use"})\, d\varphi \tag{11}$$

where $qy - c$ is the NPV of a project that receives funding and therefore the first term represents the value added of all funded projects, while the second term is total privacy costs.[47]

At the policy maker's decision stage, the state of the economy is uncertain, and therefore the policy maker's objective is to maximize expected welfare, $E[W]$, to θ. Proposition 4 derives the closed-form expression for expected welfare.

Proposition 4 *The policy maker's objective is given by*

$$\max_{\theta} E[W] = \max_{\theta} \{\gamma W_h + (1-\gamma) W_l\} \tag{12}$$

where

$$W_h = \frac{3}{4} y_h - c - 2\alpha\theta \left[1 - \left(1 + \frac{4\alpha}{3c}\right) E[q|\,u]\right] \tag{13}$$

in which the closed-form expression for $E[q|\,u]$ is in (30) and

$$W_l = \theta\lambda \tag{14}$$

where λ is a collection of constants shown in (51).

Proof. See Appendix A (p.49). ■

In the bad state, the loan market for the unrevealed is inoperative because it is dominated by bad borrowers. The choice of whether to use cash or DC for payments becomes purely individual and unaffected by other households' choices, as discussed in Section 3.2. A higher

[47] Whenever any household receives credit in equilibrium (in Propositions 1, 2, and 3) then this loan comes from the home lender and therefore the return conditional on project success is y, so that a project's expected NPV is $qy - c$.

θ now adds social value because it helps households who want to reveal their types to do so, thereby ensuring that fewer positive NPV projects remain without financing. Notably, welfare increases linearly in the bad state as θ increases.

Instead, in the good state, as described in Section 3.1, all households receive loans. This includes the bad borrowers among the pool of unrevealed, as there are enough good borrowers with whom they share that pool. Therefore, the total value added of projects is independent of θ and the DC's only impact on aggregate welfare is from privacy costs. A higher θ has a direct effect on privacy costs, through the term $-\alpha\varphi\theta$ in (1), but in addition it also has an indirect effect on privacy costs: this indirect effect derives from the revelation cascades described in Section 3.1, whereby when a fraction of high q households choose to reveal, $E\left[q|\, u\right]$ declines, which incentivizes an additional fraction of relatively high quality cash users to switch to DC use, etc.[48] Put together, direct and indirect effects on privacy costs (which can be seen, respectively, from the terms θ and $E\left[q|\, u\right]$ in (13)) imply that W_h decreases more than linearly as θ increases. There are two reasons for this. The first is that the two effects are multiplicative. The direct effect applies to all DC users linearly, but the mass of DC users simultaneously expands when θ increases due to the indirect effect. The second is that, per (36), $R_u\left(\theta\right)$ increases convexly as θ rises.[49] The intuition is that, when a given fraction of relatively high credit quality borrowers leaves the pool of the unrevealed, this shrinks the pool, so that when the next equal size fraction leaves the pool, it constitutes a larger percentage of the pool and has a larger revelatory impact on those remaining in the pool.[50]

[48] Interpreted from the angle of Figure 2, the distance between the dashed line and the unbroken line becomes larger as θ increases. The indirect effect shifts the unbroken line to the left, which is a welfare negative because agents between the dashed and unbroken line would have preferred to remain unrevealed in cash-world but are led to revelation by a Lemons' Market effect.

[49] Similarly, $E\left[q|\, u\right]$ decreases convexly as θ rises. See the discussion in the paragraph below (30).

[50] A simplified model example helps clarify this: consider an unrevealed pool with 5 borrowers who always repay and 5 borrowers that never repay, while lenders have access to costless funding ($c = 1$) and engage in perfect competition. The breakeven interest rate on loans to the unrevealed pool is then $R_u = 2$. When one good borrower leaves the pool, R_u rises to $\frac{9}{4} = 2.25$. When the second good borrower leaves, R_u rises to $\frac{8}{3} \approx 2.67$. That is, the second good borrower leaving increases R_u by more than the first good borrower leaving, because when the second borrower leaves the pool is smaller.

These effects lay the foundation for an interior solution for socially optimal policy.[51] Proposition 5 records this result and its proof derives a unique closed-form solution for $\theta^* \in (0, 1)$.

Proposition 5 *When $\gamma = 1$ and the good state is certain, privacy costs dominate and the policy maker bans the DC, $\theta^* = 0$. When $\gamma = 0$ and the bad state is certain, the value of being able to signal quality dominates and the policy maker opts for laissez-faire, $\theta^* = 1$. When $\gamma \in (0, 1)$, the policy maker for some parameterizations chooses the regulate the intrusiveness of the DC with $\theta^* \in (0, 1)$, the closed-form expression for which is provided in (54).*

Proof. See Appendix A (p.51). ■

4.2 Distributional effects

The distributional effects from DC design also differ between the two states of y. In the bad state, revelation increases the expected net benefit of the revealing households without reducing that of the households that opt for cash, because the latter do not receive loans. Moreover, aggregate welfare also rises through higher corporate profits, when more households become revealed.

In the good state, high credit quality and low privacy cost households gain from the increased ability to reveal. However, different to the bad state, their gain from lower customized loan rates is at the expense of a higher loan rate for unrevealed households. Given higher privacy costs (and the aforementioned indirect effects), only a segment of revealed households with the highest credit quality and lowest privacy costs experience net gains from a more intrusive DC in the good state, while other households experience net losses. In the good state, total corporate profits are θ-invariant: a higher θ transfers profits within the corporate sector, as the DC issuer gains data access fee revenues at lenders' expense.

[51] Underlying this is the fact that $\frac{\partial W_l}{\partial \theta} > 0$ and $\frac{\partial^2 W_l}{\partial \theta^2} = 0$, while $\frac{\partial W_h}{\partial \theta} < 0$ and $\frac{\partial^2 W_h}{\partial \theta^2} < 0$, implying a concave tradeoff for the policy maker from facing an ex-ante mixture of the two states.

4.3 Policy insights

Proposition 6 next analyzes how the interior solution in Proposition 5 is affected by the distribution of project returns. Appendix B [*to be done*] considers a full distribution of states, but here we first derive the main intuitions with a simpler setup based on binary states.[52] We let $\gamma = \frac{1}{2}$ and define $\overline{y} = \frac{y_h - y_l}{2}$ where moreover $y_h = \overline{y} + s$ and $y_l = \overline{y} - s$ so that $\overline{y}$ equals the mean of the distribution and s represents the spread around the mean. Based on Proposition 6, we then discuss comparative statics to $\overline{y}$ and s. This is of particular interest from a policy perspective, because the relative returns and riskiness of borrowers are comparable characteristics across economies (and also vary per country over time), laying the foundation for a link between the optimal DC regulation in our framework and policy insights discussed earlier (p.5).

For instance, we can consider a fixed s and increase $\overline{y}$ from a low level to a high level. With $\overline{y}$ at a low level, there are too many bad borrowers when returns bottom ($y = \overline{y} - s$) and the loan market for the unrevealed shuts down. The social value of the opportunity to disclose that the DC affords good borrowers dominates aggregate welfare considerations. With the unrevealed market shut, privacy is not a public good, and individual decisions to disclose are necessarily socially optimal because they raise individual benefit, as well as corporate profits. Then, a marginally higher $\overline{y}$ means that there are more good borrowers with a desire to disclose when returns bottom, which increases the social value of revelation, and therefore socially optimal regulation softens (higher θ^*) as the gap between privately and socially optimal DC policy shrinks: $\frac{\partial \theta^*}{\partial \overline{y}} > 0$. In contrast, when $\overline{y}$ increases beyond a threshold such that the unrevealed loan market is always open, then privacy becomes a public good and households' disclosure decisions begin to impose negative externalities. At the same time, a higher θ creates no social value because lower individual loan rates to disclosing households are, in aggregate, offset by higher loan rates on the unrevealed. Therefore, $\theta^* = 0$, which

[52] One advantage of this simplicity is that it facilitates the derivation of analytical solutions. Appendix B [to be done] instead applies numerical techniques.

in turn implies a nonlinear response of θ^* to $\overline{y}$: θ^* rises as $\overline{y}$ increases, until a threshold is reached where θ^* drops to 0.[53]

Similar effects dictate the response of θ^* to s: starting from a high $\overline{y}$ and low s, a social planner would ban the DC. As the mean-preserving spread increases, the bottom side of the distribution pushes closer to territory where the unrevealed market closes. Where the threshold is crossed, the value of disclosure is greatest, because while the unrevealed market is closed, there are relatively many good borrowers who would benefit from differentiating themselves. Hence, θ^* peaks at that point and as s rises further, optimal DC regulation is tightened (lower θ^*).

Proposition 6 *If unrevealed households cannot obtain loans when $y = \overline{y} - s$, then $\frac{\partial \theta^*}{\partial \overline{y}} > 0$ and $\frac{\partial \theta^*}{\partial s} < 0$. Instead, if unrevealed households can obtain loans when $y = \overline{y} - s$, then $\theta^* = 0$ (and consequently) $\frac{\partial \theta^*}{\partial \overline{y}} = 0$, and $\frac{\partial \theta^*}{\partial s} = 0$.*

Proof. See Appendix A (p.52). ■

5 Conclusion

New technologies are fundamentally changing the way in which payments and credit provision take place, and create an unprecedented scope for data monopolization. BigTechs, in particular, seem poised to expand upon their already vast trove of knowledge about their customers and to ponder the creation of DC payment ecosystems that enhance the network effects of their platforms. Such developments can bring both unique opportunities and risks to consumers and borrowers, and policy institutions are actively grappling with the extent to which they should step in and regulate.

This paper builds a framework that is aimed at conceptualizing key tradeoffs and considerations in the regulation of a DC. The model's richness, with three types of private agents,

[53] With the generalized state distribution in Appendix B, the response of θ^* to $\overline{y}$ is non-monotonic but smooth [*to be confirmed*]: the jumps in socially optimal policy emanate from the binary state setup.

various forms of heterogeneity within the agent categories, and extensive interactions among agent decisions both within and across agent categories, underlies its ability to go beyond a binary policy discussion and speak to the degree to which private and socially optimal policies diverge, and therefore the extent to which regulation should step in.

Privately optimal design diverges from the social optimum, because the DC issuer uses its pole position as a data monopolist to play out both households and lenders against each other. The DC issuer squeezes households into a partially unraveling Lemons' Market for credit quality data, whereby the willingness of some high quality households to disclose convinces more tranches of households to differentiate themselves from the pool of unrevealed households, despite the privacy costs involved with DC use. To maximize the value of household data in its hands, the DC issuer makes the DC as intrusive as possible, which plays optimally on these Lemons' Market dynamics. With its data chest in hand, the DC issuer then aims to market this data most effectively towards lenders. Surprisingly, the DC issuer hands out the data for free to most lenders. This strategy maximizes the pressure on home lenders, whose loans ensure household projects attain their full potential, to pay up or risk losing their highest quality borrowers.

From a social planner's view, lenders' losses are the DC issuer's gain, but the same cannot be said for the impact on households. Socially optimal design aims at shielding households from being played out against each other. The challenge in doing so, is that regulation is a long-term policy, while the benefits from counteracting the intrusiveness of the DC are state dependent. In a bad state, where projects become less profitable, credit provision to unrevealed borrowers can freeze. DC intrusiveness can then be net positive, not just to a subset of households, but to aggregate welfare. The DC becomes a means to help create data that unfreezes part of the credit market by differentiating better quality households. Because only revealed households can obtain credit, there are no disclosure externalities in this state. There are still privacy costs, but each household internalizes this cost when deciding whether to choose the DC for its payments.

To the DC issuer, household disclosure is valuable in any state of the world, but to society, differentiating data has most value during downturns. Socially optimal regulation trades off the preservation of credit enabling data gathering during downturns against the costs of households that are induced into disclosure cascades in upturns. In this sense, a fundamental friction is that DC regulatory policy is long term and cannot be made state contingent.

The model shows that the above tradeoff shifts in complex ways when the distribution of returns in states changes. This facilitates a discussion of socially optimal policy in comparison to the structure of project returns. That, in turn, can be related to broad country characteristics, such as the degree of financial development (i.e., what fraction of the population is financially excluded in periods with average economic performance) and the volatility of returns.[54]

Socially optimal policy can also be interpreted from the angle of optimal CBDC design. In this respect, our framework can provide a foundation for an analysis of competing central bank and privately issued DCs. For instance, if a central bank moves first in potentially introducing a CBDC, before a BigTech introduces its DC, would CBDC design shift from the social optimum that emerges absent such competition? If so, in what ways? Would fixed costs or network effects in payment systems affect these considerations? While an exploration of these questions lies beyond the scope of this paper, they may not lie beyond the scope of the framework it provides.

[54] The volatility of returns could relate to, e.g., the prevalence of financial boom-bust cycles or the industrial structure of the economy, such as a heavy reliance on one or a few price volatilite sectors, like primary commodities.

References

Acquisti, A., Taylor, C., and Wagman, L. (2016). The Economics of Privacy. *Journal of Economic Literature*, 54(2):442–92.

Adrian, T. (2021). BigTech in Financial Services. Speech delivered to the European Parliament FinTech Working Group on June 16, 2021.

Agarwal, A. and Assenova, V. A. (2022). Mobile Money as a Steppingstone: Addressing Voids in Credit-Market Institutions through Digital Platforms. mimeo.

Agur, I., Ari, A., and Dell'Ariccia, G. (2022). Designing Central Bank Digital Currencies. *Journal of Monetary Economics*, 125:62–79.

Ahnert, T., Hoffmann, P., and Monnet, C. (2022). The Digital Economy, Privacy, and CBDC. ECB Working Papers 2662, European Central Bank.

Akerlof, G. A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, 84(3):488–500.

Athey, S., Catalini, C., and Tucker, C. (2017). The Digital Privacy Paradox: Small Money, Small Costs, Small Talk. NBER Working Papers 23488, National Bureau of Economic Research.

Babina, T., Buchak, G., and Gornall, W. (2022). Customer Data Access and Fintech Entry: Early Evidence from Open Banking. mimeo.

Bains, P., Sugimoto, N., and Wilson, C. (2022). BigTech in Financial Services: Regulatory Approaches and Architecture. IMF Fintech Note 2022/002, International Monetary Fund.

Beck, T., Gambacorta, L., Huang, Y., Li, Z., and Qiu, H. (2022). Big Techs, QR Code Payments and Financial Inclusion. BIS Working Papers 1011, Bank for International Settlements.

Bergemann, D., Bonatti, A., and Gan, T. (2022). The Economics of Social Data. *The RAND Journal of Economics*, 53(2):263–296.

Bijlsma, M., van der Cruijsen, C., and Jonker, N. (2021a). Not all Data are Created Equal - Data Sharing and Privacy. DNB Working Papers 728, De Nederlandsche Bank.

Bijlsma, M., van der Cruijsen, C., Jonker, N., and Reijerink, J. (2021b). What Triggers Consumer Adoption of CBDC? DNB Working Papers 709, De Nederlandsche Bank.

Blume, A. (2003). Bertrand Without Fudge. *Economics Letters*, 78(2):167–168.

Boissay, F., Ehlers, T., Gambacorta, L., and Shin, H. S. (2021). Big Techs in Finance: On the New Nexus Between Data Privacy and Competition. In Rau, R., Wardrop, R., and Zingales, L., editors, *The Palgrave Handbook of Technological Finance*, Springer Books, pages 855–875. Springer.

Borgonovo, E., Caselli, S., Cillo, A., Masciandaro, D., and Rabitti, G. (2021). Money, Privacy, Anonymity: What Do Experiments Tell Us? *Journal of Financial Stability*, 56:100934.

Broecker, T. (1990). Credit-Worthiness Tests and Interbank Competition. *Econometrica*, 58(2):429–52.

Brunnermeier, M. and Payne, J. (2022). Platforms, Tokens, and Interoperability. mimeo.

Brunnermeier, M. K., James, H., and Landau, J.-P. (2021). The Digitalization of Money. BIS Working Papers 941, Bank for International Settlements.

Canayaz, M., Kantorovitch, I., and Mihet, R. (2021). Privacy Laws and Value of Personal Data. Research Paper Series 21-92, Swiss Finance Institute.

Charlson, G. (2021). Third-Degree Price Discrimination in the Age of Big Data. Working Papers in Economics 2159, University of Cambridge.

Chiu, J. and Koeppl, T. V. (2022). PayTech and the D(ata) N(etwork) A(ctivities) of BigTech Platforms. Queen's Economics Department Working Paper 1490, Queen's University.

Choi, J. P., Jeon, D.-S., and Kim, B.-C. (2019). Privacy and Personal Data Collection with Information Externalities. *Journal of Public Economics*, 173:113–124.

Cicala, S., Hémous, D., and Olsen, M. (2021). Adverse Selection as a Policy Instrument: Unraveling Climate Change. mimeo.

Conitzer, V. (2016). On Stackelberg Mixed Strategies. *Synthese*, 193:689–703.

Cornelli, G., Frost, J., Gambacorta, L., Rau, R., Wardrop, R., and Ziegler, T. (2020). Fintech and Big Tech Credit: A New Database. BIS Working Papers 887, Bank for International Settlements.

Croxson, K., Frost, J., Gambacorta, L., and Valletti, T. (2022). Platform-Based Business Models and Financial Inclusion. BIS Working Papers 986, Bank for International Settlements.

Di Maggio, M., Ratnadiwakara, D., and Carmichael, D. (2021). Invisible Primes: Fintech Lending with Alternative Data. mimeo.

Dorfleitner, G., Hornuf, L., and Kreppmeier, J. (2021). Promise not Fulfilled: FinTech Data Privacy, and the GDPR. CESifo Working Paper Series 9359, CESifo.

Fishman, M. J., Parker, J. A., and Straub, L. (2020). A Dynamic Theory of Lending Standards. Working Paper 27610, National Bureau of Economic Research.

Frost, J., Gambacorta, L., Huang, Y., Shin, H. S., and Zbinden, P. (2020). BigTech and the Changing Structure of Financial Intermediation. *Economic Policy*, 34(100):761–799.

FSB (2021). Regulation, Supervision and Oversight of "Global Stablecoin" Arrangements: Progress Report on the Implementation of the FSB High-Level Recommendations. Report Oct. 7, 2021, Financial Stability Board.

FSB (2022). Assessment of Risks to Financial Stability from Crypto-Assets. Report Feb. 16, 2022, Financial Stability Board.

Gambacorta, L., Khalil, F., and Parigi, B. M. (2022). Big Techs vs Banks. mimeo.

Garratt, R. and Lee, M. J. (2021). Monetizing Privacy. Staff Reports 958, Federal Reserve Bank of New York.

Garratt, R., Lee, M. J., Martin, A., and Townsend, R. M. (2021). Rent-Seeking Platforms. Staff Reports 892, Federal Reserve Bank of New York.

Garratt, R. J. and Van Oordt, M. R. (2021). Privacy as a Public Good: A Case for Electronic Cash. *Journal of Political Economy*, 129(7):2157–2180.

Ghosh, P., Vallee, B., and Zeng, Y. (2021). FinTech Lending and Cashless Payments. mimeo.

Grossman, S. J. (1981). The Informational Role of Warranties and Private Disclosure about Product Quality. *The Journal of Law and Economics*, 24(3):461–483.

Grossman, S. J. and Hart, O. D. (1980). Disclosure Laws and Takeover Bids. *The Journal of Finance*, 35(2):323–334.

He, Z., Huang, J., and Zhou, J. (2022). Open Banking: Credit Market Competition When Borrowers Own the Data. *Journal of Financial Economics - forthcoming*.

Hu, B., Liu, Y.-L., and Yan, W. (2021). Should I Scan my Face? The Influence of Perceived Value and Trust on Chinese Users Intention to Use Facial Recognition Payment. 23rd biennial conference, International Telecommunications Society.

Huang, J. (2021). Fintech Expansion. mimeo.

Huang, Y., Zhang, L., Li, Z., Qiu, H., Sun, T., and Wang, X. (2020). Fintech Credit Risk Assessment for SMEs: Evidence from China. IMF Working Paper 20/193, International Monetary Fund.

Ichihashi, S. (2022). Dynamic Privacy Choices. Staff Working Papers 22-08, Bank of Canada.

IMF (2020). Digital Money Across Borders: Macro-Financial Implications. Policy Paper 2020/050, International Monetary Fund.

IMF (2021). Global Financial Stability Report, Oct. 2021. Chapter 2, International Monetary Fund.

IMF (2022). Elements of Effective Crypto Policies. Policy Paper (forthcoming), International Monetary Fund.

Jin, G. Z., Luca, M., and Martin, D. (2021). Is No News (Perceived As) Bad News? An Experimental Investigation of Information Disclosure. *American Economic Journal: Microeconomics*, 13(2):141–73.

Jin, Y. and Vasserman, S. (2021). Buying Data from Consumers: The Impact of Monitoring Programs in U.S. Auto Insurance. Working Papers 29096, National Bureau of Economic Research.

Jovanovic, B. (1982). Truthful Disclosure of Information. *The Bell Journal of Economics*, 13(1):36–44.

Kahn, C. M., McAndrews, J., and Roberds, W. (2005). Money is Privacy. *International Economic Review*, 46(2):377–399.

Kocherlakota, N. R. (1998). Money Is Memory. *Journal of Economic Theory*, 81(2):232–251.

Markovich, S. and Yehezkel, Y. (2021). Data Regulation: Who Should Control Our Data? mimeo.

Mester, L. J. (2020). Fintech, Bigtech, and Financial Inclusion. Speech delivered at the Fourth Annual Fintech Conference, Federal Reserve Bank of Philadelphia, November 9, 2020.

Milgrom, P. R. (1981). Good News and Bad News: Representation Theorems and Applications. *The Bell Journal of Economics*, 12(2):380–91.

Milgrom, P. R. (2008). What the Seller Won't Tell You: Persuasion and Disclosure in Markets. *Journal of Economic Perspectives*, 22(2):115–131.

Ouyang, S. (2022). Cashless Payment and Financial Inclusion. mimeo.

Parlour, C. A., Rajan, U., and Walden, J. (2022a). Payment System Externalities. *Journal of Finance*, 77(2):1019–1053.

Parlour, C. A., Rajan, U., and Zhu, H. (2022b). When FinTech Competes for Payment Flows. *Review of Financial Studies - forthcoming*, 35(11):4985–5024.

Piolatto, A. and Schuett, F. (2022). Information vs Competition: How Platform Design Affects Profits and Surplus. CentER Discussion Paper 2022-005, Tilburg University.

Škrinjarić, B., Budak, J., and Rajh, E. (2022). Change of Consumers' Attitudes in Response to an Online Privacy Violation Incident. EIZ Working Papers 2202, The Institute of Economics, Zagreb.

Uno, Y., Sonoda, A., and Bessho, M. (2021). The Economics of Privacy: A Primer Especially for Policymakers. Working Papers 21-E-11, Bank of Japan.

6 Appendix

A Proofs

This appendix contains the proofs of the Propositions in the main text. All calculations are performed in a *Mathematica* file that is available on request from the authors.

Proof of Proposition 1. The proof proceeds by backward induction. We first note that equations (8) and (9) are considered the limit cases of, respectively, $R_k(q) = \frac{c}{q} + \tau - \varepsilon$ with $\varepsilon \to 0$ and $R_u = \frac{c}{E[q|u]} + \tau - \varepsilon$ with $\varepsilon \to 0$.[55] Moreover, to save on notation below, we focus on a single subset of the households with a given home lender (as the results extend symmetrically) and let h refer to the home lender, while using i for the other lenders. Thus, $R_{kh}(q)$ and $R_{ki}(q)$ are the differentiated loan rates that, respectively, the home lender and other lenders charge revealed households (among the subset of households for who h is the home lender); and R_{uh} and R_{ui} are the loan rates charged to all unrevealed households by, respectively, the home lender and other lenders. We also note that the proof focuses on finding the Subgame Perfect Nash Equilibrium (SPNE) in symmetric pure strategies, as pertains to the play among the other lenders.[56]

Stage 6. This is the last decision stage of the game outlined in Section 2.5. At Stage 6, some households' credit qualities are unrevealed (namely, those that chose cash in Stage 2 and those that chose DC in Stage 2 but remained unrevealed at Stage 3), while others are revealed (namely, those that chose the DC in Stage 2 and, with probability θ, became revealed at Stage 3). If a household does not borrow at Stage 6, then its profitable opportunity is lost and it makes 0 profit. If a household borrows from the home lender, then the profit on a project, if successful at Stage 7, is $y - R_{uh}$ and $y - R_{kh}(q)$, respectively, if the household is unrevealed and if it is revealed, and 0 if the project is unsuccessful. If borrowing from another lender, these expressions are $y - \tau - R_{ui}$ and $y - \tau - R_{ki}(q)$. The Stage 6 optimal choice can therefore be summarized by the following decision algorithm. If unrevealed at Stage 3, borrow from the home lender if $y - R_{uh} > y - \tau - R_{ui}$ and $y - R_{uh} > 0$; borrow from another lender if $y - \tau - R_{ui} > y - R_{uh}$ and $y - \tau - R_{uh} > 0$; do not borrow if $y - R_{uh} < 0$ and $y - \tau - R_{ui} < 0$. If revealed at Stage 3, borrow from the home lender if $y - R_{kh}(q) > y - \tau - R_{ki}(q)$ and $y - R_{kh}(q) > 0$; borrow from another lender if $y - \tau - R_{ki}(q) > y - R_{kh}(q)$ and $y - \tau - R_{ki}(q) > 0$; do not borrow if $y - R_{kh}(q) < 0$ and $y - \tau - R_{ki}(q) < 0$.

Stage 5. Lenders' breakeven loan rates are $\frac{c}{q}$ for revealed households and $\frac{c}{E[q|u]}$ for the unrevealed.[57] Due to Bertrand competition, loan rates among other lenders are given by the breakeven rates. Implicit in $R_{ki}(q) = \frac{c}{q}$ is that, if any other lenders have access to households' credit quality data, then there is always more than one such other lender with

[55] As previously noted, we apply a breakeven preference in favor of the home lender to ensure that ε can be taken to 0 in equilibrium.

[56] However, this is shown to directly extend to a mixed strategy equilibrium if N is large enough.

[57] Strictly, these are *expected* breakeven loan rates, given the uncertain realizations of project success at Stage 7. However, with a continuum of borrowers, this distinction is immaterial.

access to household data (i.e., other lenders are never data access monopolists at Stage 5), as follows from symmetric pure strategies given $N \geq 3$ at Stage 4.[58]

For unrevealed households, it now follows directly that the home lender sets the loan rate as per (9) and that they all choose to borrow from the home lender. By setting $R_{u1} = \frac{c}{E[q|u]} + \tau - \varepsilon$ with $\varepsilon \to 0$, the home lender ensures that unrevealed borrowers prefer its loan over other lenders at Stage 6, since $(y - R_{uh}) - (y - \tau - R_{ui}) = -\varepsilon$. Moreover, all unrevealed households choose to borrow, since $y - R_{uh} = y - \frac{c}{E[q|u]} - \tau + \varepsilon$ and from $y > 2c + \tau$ and $q \in \left[\frac{1}{2}, 1\right]$ we have that $\inf \left(y - \tau - \frac{c}{E[q|u]} + \varepsilon\right) = \varepsilon$.[59]

For revealed households, we need to identify separate cases that depend on lender data purchase decisions at Stage 4: 1) if both the home lender and other lenders obtained the data at Stage 4, then, following the same limit pricing argument as in the previous paragraph, the home lender sets R_{kh} in accordance with (8) and all revealed households choose to borrow from it rather than borrowing from other lenders or not borrowing;[60] 2) if neither the home lender nor other lenders obtained the data at Stage 4, then, household revelation at Stage 3 is moot and all households are effectively unrevealed; 3) if the home lender purchases the data at Stage 4 while the other lenders do not, then the home lender optimally disregards the data that it has purchased and simply charges R_{uh} according to (9) to all borrowers, because they are "captive" at that rate (cannot do better at other lenders, because they are unrevealed to other lenders) and there is therefore no incentive for the home lender to offer higher quality borrowers customized (lower) loan rates; 4) the case where the home lender did not purchase the data in Stage 4 while other lenders did, is more intricate. We analyze this case as part of Stage 4 and refer to it as the "deviation" case. That is, the conjectured equilibrium is as given by Proposition 1 and this is sustainable as an equilibrium if the deviation case can be excluded.

Stage 4. Other lenders only obtain household data if the DC issuer offers them free access. This follows from the loan rate setting at Stage 5: other lenders cannot make a positive profit on their loan portfolio, whether they have access to the data or not, and therefore will not purchase the data at a positive price. Our focus here is on symmetric pure strategy play among the other lenders, where given $N \geq 3$ there are always at least two other lenders (in addition to the home lender).[61] If the DC issuer charges the other lenders a

[58] This can also from mixed strategies with N large enough.

[59] This infimum comes from the fact that $y > 2c + \tau$ and therefore $\inf \left(y - \tau - \frac{c}{E[q|u]}\right) = 2c - \frac{c}{E[q|u]}$ where $E\left[q|\, u\right] \geq \frac{1}{2}$ since $q \geq \frac{1}{2}$ and therefore $\inf \left(2c - \frac{c}{E[q|u]}\right) = 0$.

[60] That revealed households certainly choose to participate in lending again derives from $y > 2c + \tau$ and $q \in \left[\frac{1}{2}, 1\right]$. When borrowing from the home lender, a revealed household (conditional on project success) earns $y + \varepsilon - \frac{c}{q} - \tau$ where $y > 2c + \tau$ while $\frac{c}{q} < 2c$ given $q \in \left[\frac{1}{2}, 1\right]$.

[61] This can be extended to mixed strategies (among other lenders) if N is large enough. Mixed strategies would imply the assignment of a probability in $[0, 1]$ for the purchase of DC issuer data by the other lenders in Stage 4. Consider the limit case of $N \to \infty$. For any data purchase probabiliy > 0, there will be a continuum of lenders purchasing data at Stage 4 and therefore competing to limit pricing at Stage 5, meaning their willingness to pay for the data at Stage 4 remains 0. Put differently, with $N \to \infty$ the optimal probability assigned by other lenders to the purchase of data (at a positive price) is zero, since any nonzero probabilty implies an expected loss. For more on conditions for equivalence between pure and mixed strategies in sequential games, we refer to Conitzer (2016).

positive fee for data access, then given that the other lenders will not buy the data, the home lender will not be willing to pay a fee for data access either. Per point 3) in the previous paragraph, data access has no value added to the home lender if no other lenders have data access.

Therefore, the Stage 4 choice centers on whether the home lender buys the data when the other lenders have received that data for free. Providing that data for free to other lenders, is necessarily optimal for the DC issuer, as this will allow it to charge a positive fee to the home lender. In our conjectured equilibrium, the optimal data fee that the DC issuer charges the home lender equals the full profits made on revealed households. For a given revealed household, the expected profit that the home lender makes in equilibrium (per 8) is $q\left(\frac{c}{q}+\tau\right)-c=q\tau$. This means that the profit (excluding data access fees) that home lender makes on the full set of revealed households in equilibrium is given by $mE\left[q|\, dc\right]\tau$, where we recall that m is the mass of revealed borrowers and where $E\left[q|\, dc\right]$ is the expected quality of DC users (and since all revealed households are randomly drawn DC users, this is identical to the expected quality of DC users).[62] Both m and $E\left[q|\, dc\right]$ are solved in closed form further down in this proof.

The validity of the conjectured equilibrium therefore centers on the question of the deviation case: will the home lender choose to deviate from the equilibrium by not purchasing the data? If the home lender deviates, it can no longer differentiate among households. This means that it must charge all borrowers, both revealed and unrevealed, the same rate, which we will refer to as R_{dev}. At $R_{dev}=\frac{c}{E[q|u]}+\tau$, the conjectured equilibrium loan rate on the unrevealed, the home lender would fail to attract any revealed borrowers. The emanates from the fact that all revealed households have a better credit quality than the average of unrevealed households. We can show this by using expression (7) - household sorting in the conjectured equilibrium - to obtain an expression for the q of the lowest quality revealed household. Only DC users can become revealed and, thus, the lowest possible quality revealed borrower is identical to the lowest quality DC user. This user can be found from setting (7) to an equality (i.e., an indifference frontier between DC and cash use) and setting $\varphi=0$: from (7), the DC user with the lowest privacy costs, is also the DC user with the lowest credit quality. From (7) (and replacing from 9 for R_u) in the credit quality of this household is $q\left(\left(\frac{c}{E[q|u]}+\tau\right)-\tau\right)-c=0\Leftrightarrow q=E\left[q|\, u\right]$. Hence, the lowest quality revealed household has the same credit quality as the average of the pool of the unrevealed and all other revealed households therefore have $q>E\left[q|\, u\right]$. Recalling that other lenders offer revealed households $R_{ki}\left(q\right)=\frac{c}{q}$, this means that when borrowing from other lenders, a revealed household earns $y-\tau-\frac{c}{q}$ on a successful project; instead, if a revealed household were to borrow from the home lender at $R_{dev}=\frac{c}{E[q|u]}+\tau$ then it would earn $y-\left(\frac{c}{E[q|u]}+\tau\right)$ on a successful project. Here, $y-\tau-\frac{c}{q}>y-\left(\frac{c}{E[q|u]}+\tau\right)\Leftrightarrow q>E\left[q|\, u\right]$, which is true for the full mass of revealed households, m (as the weight of the threshold revealed household with $q=E\left[q|\, u\right]$ is zero).

[62] Strictly speaking, the optimal DC issuer fee is $mE\left[q|\, dc\right]\tau-\varepsilon$ with $\varepsilon\rightarrow0$. An implicit assumption could be that the home lender has a breakeven preference for buying the data, so that the equilibrium remains sustained at $\varepsilon=0$.

This means that if the home lender chooses to deviate, it must optimally do so by charging all households a loan rate $R_{dev} < \frac{c}{E[q|u]} + \tau$ at Stage 5, since it cannot improve its profits by not purchasing the data while charging the conjectured equilibrium loan rate on the unrevealed. If it were to charge $R_{dev} > \frac{c}{E[q|u]} + \tau$ then the home lender would loose all unrevealed households to other lenders, without gaining any of the revealed. Therefore, the deviation, if any, must be in the direction where the home lender accept lower returns on all the unrevealed households borrowing from it (by lowering R_{dev}), while gaining new borrowers from the pool of revealed households. The rest of the proof of Stage 4 concerns itself with showing that a strategy wherein the home lender does not purchase the data (at the conjectured equilibrium fee) and charges $R_{dev} < \frac{c}{E[q|u]} + \tau$ is suboptimal for the home lender, as compared to following the outlined equilibrium strategy. In particular, we show that $\tau < \frac{1}{3}c$ is a sufficient condition to ensure this.[63]

We first offer an economic intuition for why an upper bound on τ makes deviation suboptimal. In essence, τ compensates for the adverse selection problem that the uninformed home lender faces when all other lenders others buy the data. Consider, for example, the highest quality household with $q = 1$. When borrowing from informed other lenders, this household earns $y - \tau - c$ on its successful project. When borrowing from the deviating home lender, it earns $y - R_{dev}$. This means that it only chooses the home lender if $R_{dev} < c + \tau$. When τ is small, the home lender cannot afford to make R_{dev} low enough to attract it, as it would make a loss on its loan portfolio overall. Instead, the lowest quality revealed household, with $q = E[q|\, u]$ switches to the home lender as soon as R_{dev} declines marginally below $\frac{c}{E[q|u]} + \tau$. That is, there is an adverse selection among revealed borrowers when the home lender tries to attract them and this adverse selection problem is too large when τ is small, so that the home lender does not choose to deviate.

To prove the sufficient condition, we first require an expression for the profit of the deviating home lender:

$$\Pi_{dev} = \int_{E(q|u)}^{\frac{c}{R_{dev}-\tau}} (qR_{dev} - c) f(q) dq + (1-m)\left(E(q|u)R_{dev} - c\right) \tag{15}$$

where $\int_{E(q|u)}^{\frac{c}{R_{dev}-\tau}} (qR_{dev} - c) f(q) dq$ and $(1-m)\left(E(q|u)R_{dev} - c\right)$, respectively, represent the profit on revealed and unrevealed borrowers. The expected profit per revealed borrower is $qR_{dev} - c$ and the set of revealed households that chooses to borrow from the home lender runs from $q = E(q|u)$ to $q = \frac{c}{R_{dev}-\tau}$. The latter comes from setting to equality and solving to q the household's condition to prefer borrowing from the uninformed home lender relative to informed other lenders: $y - R_{dev} \geq y - \tau - \frac{c}{q}$. Moreover, $f(q) = \frac{c}{\alpha}\left(\frac{q}{E[q|u]} - 1\right)$, as derived from the indifference condition that determines which households opt for DC use, as derived at Stage 2. Even though Stage 2 occurs before Stage 4, the same sorting condition applies

[63] A more general approach than a sufficient condition would be to solve the first order condition of (15) to R_{dev} and find the highest profit that the deviating lender could make (for any τ). The difference between this profit and the equilibrium profit of the home lender would then be the optimal DC issuer data access fee (which would be identical to the current expression for τ small enough, but beyond a point would decline as τ increases). However, the first order condition to (15) is fourth order in R_{dev} and does not readily lend itself to analysis.

here, because we are investigating a deviation from the conjectured equilibrium. That is, backward induction to Stage 1 proceeds as if the home lender chooses not to deviate (as below) and then we check (here) that this non-deviation is indeed optimal. Lastly, the expected profit of the home lender per unrevealed household is $E(q|u)R_{dev} - c$ and therefore total profit on the mass, $1-m$, of unrevealed households is given by $(1-m)(E(q|u)R_{dev} - c)$.

A sufficient condition for non-deviation is $\frac{\partial \Pi_{dev}}{\partial R_{dev}} > 0$ for $R_{dev} \leq \frac{c}{E[q|u]} + \tau$. Combined with the fact that $R_{dev} > \frac{c}{E[q|u]} + \tau$ is necessarily suboptimal to home lender (as it leads to zero profits, due to the loss of unrevealed borrowers to other lenders), this would imply $R^*_{dev} = \frac{c}{E[q|u]} + \tau$ which is the equilibrium loan rate. The proof that $\tau < \frac{1}{3}c \Rightarrow \frac{\partial \Pi_{dev}}{\partial R_{dev}} > 0$ is provided in the next footnote, to save on space. What this derivative means is that, to the home lender, lost revenues on unrevealed households from cutting R_{dev} always outweigh the additional revenues on the additional revealed borrowers that choose the home lender when it cuts R_{dev}.[64]

Stage 2. Stage 3 is not a decision stage and therefore we next turn to Stage 2. At Stage 2, a household prefers DC over cash if and only if

$$q \max\{(y - \tau I - R_{dc}(q)), 0\} - \alpha\varphi\theta > q \max\{(y - \tau I - R_{cash}), 0\} \quad (16)$$

where $R_{dc}(q)$ is the loan rate that the household would expect if it chose to use the DC and R_{cash} is the loan rate it would receive if it chose cash. The loan rate on cash is not type dependent, because cash users do not become revealed and therefore are offered loan rate R_u (note that we here return to the notation R_u for the equilibrium loan rate on the unrevealed

[64] A sufficient condition for $\frac{\partial \Pi_{dev}}{\partial R_{dev}} > 0$ is that $\frac{\partial}{\partial R_{dev}} \left[\int_{E(q|u)}^{\frac{c}{R_{dev}-\tau}} 2(qR_{dev} - c)dq + (1-m)\left(\frac{1}{2}R_{dev} - c\right) \right] > 0$. There are two reasons this is sufficient. First, in (15), $\int_{E(q|u)}^{\frac{c}{R_{dev}-\tau}} (qR_{dev} - c)f(q)dq$ is where the potential negative part of $\frac{\partial}{\partial R_{dev}}$ comes from and this term is necessarily larger (in absolute terms) if integrated over $f(q) = 2$. This is portion where the indifference frontier (as displayed in Figure 2) has "hit" the upper bound of $\varphi = 1$; there we deal with a uniform distribution (which has $f(q) = 2$ given $q \in \left[\frac{1}{2}, 1\right]$). There, the distribution always contains more mass than if it is integrated over a portion where $f(q)$ is smaller (sloping part of indifference frontier in Figure 2). Put differently, the potential for cutting R_{dev} to raise profits comes from the fact that the home lender can entice more revealed households to borrow from it. The mass of revealed borrowers switching for a given rate cut is largest when $f(q) = 2$. Hence, for this proof, it is sufficient to consider $f(q) = 2$ rather than the more general $f(q)$. Secondly, from (15), $\frac{\partial}{\partial R_{dev}}(1-m)(E(q|u)R_{dev} - c)$ is always positive and for this term $(1-m)(E(q|u)R_{dev} - c) > (1-m)\left(\frac{1}{2}R_{dev} - c\right)$ (note that $\inf E(q|u) = \frac{1}{2}$). That is, with these steps, we can only weaken the positive effect and strengthen the negative, implying that if the overall derivative to R_{dev} is nonetheless positive, then this is sufficient for the more general case.

Solving the integral and taking the derivative the above sufficent condition becomes $\frac{R_h - 3\tau}{(R_h - \tau)^3} > \frac{2[E(q|u)]^2 + m - 1}{2c^2}$. This condition is at its tightest when $E(q|u) = \sup E(q|u) = E(q) = \frac{3}{4}$ and $m = \sup m = 1$. Replacing these, a sufficient condition is $\frac{R_{dev} - 3\tau}{(R_{dev} - \tau)^3} > \frac{9}{32c^2}$. We next turn to finding the infimum of $\frac{R_{dev} - 3\tau}{(R_{dev} - \tau)^3}$. We have that $\frac{\partial}{\partial R_{dev}} \frac{R_{dev} - 3\tau}{(R_{dev} - \tau)^3} = 0$ is $R_{dev} = 4\tau$ but that is a maximum as $\left(\left(\frac{\partial^2}{\partial R_{dev}^2} \frac{R_{dev} - 3\tau}{(R_{dev} - \tau)^3} R_{dev} = 4\tau\right) = -\frac{2}{81}\tau^4 < 0\right.$. Hence, the infimum of $\frac{R_{dev} - 3\tau}{(R_{dev} - \tau)^3}$ must lie in its corners. Investigating these corners ($R_{dev} \in \left(\frac{4}{3}c, 2c + \tau\right)$ are maximum bounds), the sufficient condition becomes $\frac{9(4c - 9\tau)}{(4c - 3\tau)^3} > \frac{9}{32c^2}$ which (given $\tau > 0$) can be solved to $\tau < \frac{c\left(8 - 4\sqrt{3}\right)}{3}$ and since $8 - 4\sqrt{3} > 1$, it suffices to set $\tau < \frac{c}{3}$.

and $R_k(q)$ for the equilibrium loan rate on revealed borrowers, as the equilibrium among the lenders is backward induced from the later stages). Instead, $R_{dc}(q) = \theta R_k(q) + (1-\theta) R_u$, because a DC user's credit quality is revealed with probability θ in which case that household receives the customized loan offer $R_k(q)$, while with probability $1-\theta$ the household remains unrevealed and pays R_u on its loan. Moreover, since every household chooses to participate in the loan market at Stage 6, the max operators in the expression above can be removed (i.e., the terms on the left of the operator are necessarily larger than 0). Therefore, we can rewrite the above inequality to:

$$q(y - \tau I - (\theta R_k(q) + (1-\theta) R_u)) - \alpha\varphi\theta > q(y - \tau I - R_u) \tag{17}$$

which can be written to

$$\alpha\varphi < q(R_u - R_k(q)) \tag{18}$$

and replacing from equation (8) this becomes (7).

Stage 1. At Stage 1, the DC issuer optimally sets $mE[q|\, dc]\,\tau$ as access fee to the home lender and provides the data for free to the other lenders. This follows from the Stage 4 derivations: the DC issuer can only sell the data at a positive price to any lender if it provides the data for free to the other lenders and charges a fee to the home lender. The highest fee it can charge the home lender, such that the home lender is still willing to purchase the data, is $mE[q|\, dc]\,\tau - \varepsilon$ with $\varepsilon \to 0$ (a higher fee would larger than the profit that the home lender can make on revealed households; and a lower fee is not needed to entice the home lender to buy the data, since at this fee the home lender does not deviate, as shown at Stage 4).

At Stage 1, the DC issuer also determines θ. Since its profits are given by $mE[q|\, dc]\,\tau$ (time N but this is irrelevant for the proof), its aim is to maximize $mE[q|\, dc]$ to θ. We derive closed form solution for m and for $E[q|\, dc]$ and show that $\frac{\partial mE[q|dc]}{\partial\theta} > 0$ and therefore $\theta = 1$ is optimal. We now write m as a function of θ and note that

$$m(\theta) = \theta\mu(\theta) \tag{19}$$

where $\mu(\theta)$ is the mass of DC users. In the SPNE, $m(\theta) = \mu(\theta)$, but this is due to the optimality of $\theta = 1$ which we must first derive. We can find the mass of DC users by integrating over the indifference frontier displayed as the unbroken line in Figure 2. As can be seen from that figure, defining the area of integration requires finding two values of q: 1) the lowest q household that is a DC user (which is the starting point of the integral); 2) the lowest value of q above which all users (regardless of φ) are DC users. From the first to the second point, we must integrate over the indifference frontier and from the second point till $q = 1$ we must integrate over mass 1. The first point we have already established at Stage 4 and is given by $q = E[q|\, u]$. The second point is derived from entering $\varphi = 2$ in the expression for the indifference frontier, which yields $2\alpha = c\left(\frac{q}{E[q|u]} - 1\right)$ and can be written to:

$$q = \frac{2\alpha + c}{c} E[q|\, u] \tag{20}$$

Here, we note that $\frac{1}{2} \leq E[q|\, u] < \frac{2\alpha+c}{c} E[q|\, u] < 1$, where $\frac{2\alpha+c}{c} E[q|\, u] < 1$ follows from the condition in (6).[65] This, in turn, ensures that the two pieces of integration shown in the equation below are properly defined, given $q \in \left[\frac{1}{2}, 1\right]$.

We can now write

$$\mu(\theta) = \int_{E[q|u]}^{\frac{2\alpha+c}{c}E[q|u]} f(q)\, dq + \int_{\frac{2\alpha+c}{c}E[q|u]}^{1} 2dq \tag{21}$$

$$f(q) = \frac{c}{\alpha}\left(\frac{q}{E[q|\, u]} - 1\right) \tag{22}$$

where $f(q)$ is the cash-DC use indifference frontier derived from setting (7) to equality and writing to φ. Moreover, the 2 in each integral in (21) comes from the uniform distribution: given that $q \in \left[\frac{1}{2}, 1\right]$, the probability density functions include a term $\frac{1}{1-1/2} = 2$, whereby $\int_{1/2}^{1} 2dq = 1$ as per the unit mass of households. We note that $E[q|\, u]$ is a function of θ (derived further below). Solving (21) gives:

$$\mu(\theta) = 2\left(1 - \frac{c+\alpha}{c}E[q|\, u]\right) \tag{23}$$

and thus

$$m(\theta) = 2\theta\left(1 - \frac{c+\alpha}{c}E[q|\, u]\right) \tag{24}$$

where we note that the term in parenthesis is always positive given $E[q|\, u] \in \left[\frac{1}{2}, \frac{3}{4}\right]$ and $\alpha < \frac{1}{6}c$ from (6).

To progress on the derivative to θ, we therefore need work towards a closed-form expression for $E[q|\, u]$. Here, we first note that

$$(1 - m(\theta))\, E[q|\, u] + m(\theta)\, E[q|\, dc] = E[q] = \frac{3}{4} \tag{25}$$

which can be written to

$$E[q|\, u] = \frac{3/4 - m(\theta)\, E[q|\, dc]}{1 - m(\theta)} \tag{26}$$

where $E[q|\, dc]$ is the expected credit quality of DC users. We next require an expression for $E[q|\, dc]$, which is a weighted average of expected values with respect to the integral areas

[65] I.e., $\sup E[q|\, u] = E[q] = \frac{3}{4}$ which implies $\sup \frac{2\alpha+c}{c}E[q|\, u] = \frac{2\alpha+c}{c}\frac{3}{4}$ and this is smaller than 1 given $\alpha < \frac{1}{6}c$ from (6).

identified in (21):

$$E\left[q|\,dc\right]=\frac{\int\limits_{E[q|u]}^{\frac{2\alpha+c}{c}E[q|u]} q\left(\frac{c}{\alpha}\left(\frac{q}{E[q|u]}-1\right)\right)dq+\int\limits_{\frac{2\alpha+c}{c}E[q|u]}^{1} 2qdq}{\int\limits_{E[q|u]}^{\frac{2\alpha+c}{c}E[q|u]} \frac{c}{\alpha}\left(\frac{q}{E[q|u]}-1\right)dq+\int\limits_{\frac{2\alpha+c}{c}E[q|u]}^{1} 2dq} \tag{27}$$

and solving this gives

$$E\left[q|\,dc\right]=\frac{4\alpha^2\left(E\left[q|\,u\right]\right)^2+3c^2\left(\left(E\left[q|\,u\right]\right)^2-1\right)+6\alpha c\left(E\left[q|\,u\right]\right)^2}{6c\left(\alpha E\left[q|\,u\right]+c\left(E\left[q|\,u\right]-1\right)\right)} \tag{28}$$

Replacing from (24) and (28) into (26) and solving leads to this expression for $E\left[q|\,u\right]$.

$$E\left[q|\,u\right]=\frac{16\alpha^2\theta\left(E\left[q|\,u\right]\right)^2+3c^2\left(3+4\theta\left(\left(E\left[q|\,u\right]\right)^2-1\right)\right)+24\alpha c\theta\left(E\left[q|\,u\right]\right)^2}{12c\left(2\alpha\theta E\left[q|\,u\right]+2c\theta\left(E\left[q|\,u\right]-1\right)+c\right)} \tag{29}$$

We can solve this in closed form for $E\left[q|\,u\right]$:

$$E\left[q|\,u\right]=\frac{1}{\theta}\frac{3c^2\left(2\theta-1\right)+\sqrt{3}\sqrt{3c^4\left(1-\theta\right)+4\alpha^2c^2\theta\left(4\theta-3\right)}}{6c^2+8\alpha^2} \tag{30}$$

where an additional negative root solution can be discarded as it always gives $E\left[q|\,u\right]<\frac{1}{2}$, which violates $q\in\left[\frac{1}{2},1\right]$.[66] It can also be verified that for the solution in (30), it holds that $E\left[q|\,u\right]\in\left[\frac{1}{2},\frac{3}{4}\right]$.[67] We also note that the apparent singularity at $\theta=0$ is not a genuine singularity. Its appearance comes from the fact that (30) is the solution of (29) when (29) is written as a quadratic equation. However, at $\theta=0$, (29) simplifies to $E\left[q|\,u\right]=\frac{3c^2(3)}{12c(c)}=\frac{3}{4}$ and there is no quadratic equation to solve: when $\theta=0$, all households are necessary unrevealed and $E\left[q|\,u\right]=E\left[q\right]=\frac{3}{4}$. Furthermore, we note that $\frac{\partial E[q|u]}{\partial\theta}<0$ and $\frac{\partial^2 E[q|u]}{\partial\theta^2}<0$.[68] As the extent of revelation of relatively high quality DC users increases, the expected quality of the unrevealed pool declines, and foreseeing this causes more households to sort into DC use in the first place: this feedback effect strengthens as θ increases further (per $\frac{\partial^2 E[q|u]}{\partial\theta^2}<0$).

We can now replace from (30) into (24) to obtain

$$m\left(\theta\right)=\frac{1}{c}\frac{3c^2\left(c+3\alpha\left(1-2\theta\right)\right)-8\alpha^2c\theta-\left(\alpha+c\right)\sqrt{3}\sqrt{3c^4\left(1-\theta\right)+4\alpha^2c^2\theta\left(4\theta-3\right)}}{3c^2-4\alpha^2} \tag{31}$$

[66] This can be seen from the fact that the negative root solution (replacing the + in the numerator of (30) with a $-$) is necessarily smaller than $\frac{2\theta-1}{\theta}\frac{3c^2}{6c^2+8\alpha^2}$ and $\frac{2\theta-1}{\theta}\frac{3c^2}{6c^2+8\alpha^2}<\frac{2\theta-1}{\theta}\frac{3c^2}{6c^2}=1-\frac{1}{2}\frac{1}{\theta}<\frac{1}{2}$.

[67] Using the $MaxValue$ and $MinValue$ functions in *Mathematica* on (30) over the allowed parameter space, $\theta\in[0,1], c\geq 1, \alpha\in\left[0,\frac{c}{6}\right]$, we obtain $\frac{3}{4}$ as the supremum of $E\left[q|\,u\right]$ and $\frac{1}{2}$ as the infimum.

[68] Using the $MaxValue$ function in *Mathematica*, we verify that $\sup_{\theta\in(0,1],c\geq1,\alpha\in\left[0,\frac{c}{6}\right]}\frac{\partial E[q|u]}{\partial\theta}<0$.

where we verify that $m(\theta) \in [0,1]$ over our parameter space.[69] Obviously, $m(\theta) = 0$ when $\theta = 0$, but is interesting to note that the possibility of $m(\theta) = 1$ also exists: when $\alpha = 0$, implying there are no privacy preferences at play, all households choose to use DC; only the presence of privacy preferences prevents a cascade to full revelation.[70]

This provides us with the elements needed to show that $\frac{\partial m(\theta)E[q|dc]\tau}{\partial\theta} > 0$ (the revenue from data selling increases as θ rises) and therefore $\theta = 1$ is optimal. In particular, replacing for $m(\theta)$ from (31) and for $E[q|\,dc]$ from (28) and (30) we obtain the closed form for the optimal data selling fee of the DC issuer to the home lender, which we denote by $\widehat{\Omega}$:

$$\begin{aligned}\widehat{\Omega} &= m(\theta)\, E[q|\,dc]\,\tau \qquad (32)\\ &= \frac{\tau}{4\theta\left(3c^2-4\alpha^2\right)^2}\begin{pmatrix} 9c^4(5\theta-2) - 18\alpha c^3(2+\theta(4\theta-5)) + 48\alpha^4\theta \\ +24\alpha^3 c\theta(3-4\theta) + 24\alpha^2c^2\left(1-4\theta+8\theta^2\right) \\ +\left(8\alpha^2+6c^2+12\alpha c\right)(1-2\theta) \\ \sqrt{3}\sqrt{3c^4(1-\theta)+4\alpha^2c^2\theta(4\theta-3)} \end{pmatrix} \qquad (33)\end{aligned}$$

and using (33) we can confirm that $\frac{\partial\widehat{\Omega}}{\partial\theta} > 0$.[71]

It is useful to also record $R_u(\theta)$ and its derivatives. From replacing (30) into (9) and simplifying, we obtain

$$R_u(\theta) = \theta c\frac{6c^2+8\alpha^2}{3c^2(2\theta-1)+\sqrt{3}\sqrt{3c^4(1-\theta)+4\alpha^2c^2\theta(4\theta-3)}} + \tau \qquad (34)$$

where[72]

$$\frac{\partial R_u(\theta)}{\partial\theta} > 0; \frac{\partial R_u(\theta)}{\partial\alpha} < 0; \frac{\partial R_u(\theta)}{\partial c} > 1; \frac{\partial R_u(\theta)}{\partial\tau} = 1 \qquad (35)$$

and using the same methodology as described in the last footnote, we also find that

$$\frac{\partial^2 R_u(\theta)}{\partial\theta^2} > 0 \qquad (36)$$

These comparative statics are intuitive: a higher θ worsens the expected quality of the pool of the unrevealed, implying a higher $R_u(\theta)$; cascade effects become stronger as θ rises further, leading to convexity in the derivative to θ; a higher α means that households care more about privacy, which improves the quality of the unrevealed pool as there are more relatively good quality households who choose to use cash and this reduces $R_u(\theta)$; higher lender funding costs, c, translate into higher loan rates; and, lastly, greater home lender market power from τ translates into higher loan rates as well. ■

[69] Using the *MinValue* and *MaxValue* functions in *Mathematica*, we find the infimum and supremum of (31) as $\inf_{\theta\in[0,1],c\geq1,\alpha\in\left[0,\frac{c}{6}\right]} m(\theta) = 0$ and $\inf_{\theta\in[0,1],c\geq1,\alpha\in\left[0,\frac{c}{6}\right]} m(\theta) = 1$.

[70] Full revelation of all households also requires $\theta = 1$, but sorting into DC use is complete ($\mu(\theta) = 1$) when $\alpha = 0$.

[71] Using the *MinValue* function in *Mathematica*, we find the infimum of $\frac{\partial\Omega^*}{\partial\theta} > 0$ as $\inf_{\theta\in[0,1],c\geq1,\alpha\in\left[0,\frac{c}{6}\right]} \frac{\partial\Omega^*}{\partial\theta} = 0$ (and strictly positive when $\theta > 0$).

[72] In particular, it is immediate that $\frac{\partial R_u(\theta)}{\partial\tau} = 1$, while for the other derivatives we use the *MinValue* function and *MaxValue* functions in *Mathematica* to calculate infima and suprema over the allowable paramater range $\theta \in [0,1], c \geq 1, \alpha \in \left[0, \frac{c}{6}\right]$.

Proof of Proposition 2. When $y < \frac{4}{3}c$, the breakeven interest rate on lending to unrevealed households is always lower than y, which means there exists no R_u at which unrevealed households would be willing to borrow. In particular, the breakeven rate on lending to unrevealed households is $\frac{c}{E[q|u]}$ and $\sup E[q|u] = E[q] = \frac{3}{4}$ so that $\inf \frac{c}{E[q|u]} = \frac{4}{3}c > y$. This considerably simplifies the problem compared to Proposition 1.

Stage 6. If a revealed household borrows from home lender or other lenders then the profit on a project, if successful at Stage 7, is $y - R_{kh}(q)$ and $y - \tau - R_{ki}(q)$, respectively. The Stage 6 optimal choice can therefore be summarized by the following decision algorithm. If revealed at Stage 3, borrow from the home lender if $y - R_{kh}(q) > y - \tau - R_{ki}(q)$ and $y - R_{kh}(q) > 0$; borrow from another lender if $y - \tau - R_{ki}(q) > y - R_{kh}(q)$ and $y - \tau - R_{ki}(q) > 0$; do not borrow if $y - R_{kh}(q) < 0$ and $y - \tau - R_{ki}(q) < 0$.

Stage 5. Lenders' breakeven loan rate for a revealed borrower of quality q is $\frac{c}{q}$. Depending on a household's quality q, we can now distinguish three cases. First, if $q < \frac{c}{y}$, then no lender can make a loan rate offer that is acceptable to the household, since $y - \frac{c}{q} < 0$. Second, if $q \in \left[\frac{c}{y}, \frac{c}{y-\tau}\right]$, then only the home lender can make a loan offer that this household could accept, because $y - \frac{c}{q} > 0$ but $y - \tau - \frac{c}{q} < 0$. Third, if $q > \frac{c}{y-\tau}$ then all lenders can make such offers since $y - \tau - \frac{c}{q} > 0$. For households with $q > \frac{c}{y-\tau}$, then, the same arguments as in the Proof of Proposition 1 continue to apply to this stage and therefore the equilibrium loan rate is given by (8) and this is the offer that the household will accept (from the home lender) at Stage 6. Instead, for households with $q \in \left[\frac{c}{y}, \frac{c}{y-\tau}\right]$, the home lender is a monopolist and therefore sets $R_{kh}(q) = y - \varepsilon$ with $\varepsilon \to 0$ to fully appropriate the return on the project (if successful).

One consideration is whether there are any households for which $q > \frac{c}{y-\tau}$. That is, by (4), we have that $q \to 1 \Rightarrow q > \frac{c}{y}$ and therefore we can be certain that there are households in the second group. But as concerns the third group, with $q > \frac{c}{y-\tau}$, this is not directly apparent. Here, for the best quality, $q = 1$, household, the condition becomes $1 > \frac{c}{y-\tau} \Leftrightarrow \tau < y - c$. Hence, the third group ($q > \frac{c}{y-\tau}$) of households exists if and only if $\tau < y - c$.

Stage 4. The deviation case where the home lender refuses to purchase the data can be straightforwardly excluded here. The reason is that any attempt of an uninformed home lender to set R_{dev} in a way that attracts revealed households, immediately induces the full set of unrevealed households to borrow from the home lender too, including many bad borrowers. Specifically, $R_{dev} < y$ is needed in order to attract any revealed households. But given limited liability, all unrevealed households that are shut out of lending in the conjectured equilibrium, will now choose borrow from the home lender too, because conditional on project success they earn $y - R_{dev}$. The home lender cannot distinguish them from revealed households. This necessarily implies the home lender makes a loss, because even if all revealed households joined the unrevealed in borrowing from the home lender, the loan portfolio would earn $E[q] R_{dev} - c = \frac{3}{4}R_{dev} - c$ which is negative since $R_{dev} < y < \frac{4}{3}c$.

Having excluded the deviation case, the same arguments as in the Proof of Proposition 1 imply that the DC issuer will offer data access for free to the other lenders and will charge the home lender a fee that appropriates all its profits on revealed households. As before, this is given by $m(\theta) E[q|\, dc]\tau$, which can be represented in closed form below.

Stage 2. Stage 3 is not a decision stage and therefore we next turn to Stage 2. Households with $q < \frac{c}{y}$ cannot obtains loans at Stage 5 and thus see only a privacy cost to using the DC. Hence, they choose cash. Households with $q \in \left[\frac{c}{y}, \frac{c}{y-\tau}\right]$, similarly see only a privacy cost to using the DC. They foresee that, if they use DC, become revealed and get a loan offer, that offer will be $R_{kh}(q) = y$ and leave no profit for them from a successful project. Thus, they too choose to use cash. If $\tau > y - c$ and there all households therefore have $q < \frac{c}{y-\tau}$, then all households choose cash and no credit provision takes place in the model. Instead, if $\tau < y - c$ and there are households with $q > \frac{c}{y-\tau}$, then at Stage 2, such a household prefers DC over cash if and only if

$$q \max\{(y - \tau I - R_{dc}(q)), 0\} - \alpha\varphi\theta > 0 \tag{37}$$

where the right hand side represents the zero returns on choosing cash, while the left-hand side represents the return when choosing the DC. This can be written to $\theta q\left(y - \left(\frac{c}{q} + \tau\right)\right) - \alpha\varphi\theta > 0$, because the unrevealed do not receive loans and with probability θ the DC user of sufficient quality becomes revealed and borrows from the home lender at the loan rate given by (8) at Stage 6. From here we have

$$\alpha\varphi < q(y - \tau) - c \tag{38}$$

where we note that (38) can also be written as $q > \frac{\alpha\varphi + c}{y-\tau}$. Since $\alpha\varphi \geq 0$, we have that $\frac{\alpha\varphi+c}{y-\tau} \geq \frac{c}{y-\tau}$ and therefore the condition in (38) is tighter than the condition $q > \frac{c}{y-\tau}$. I.e., among the households with $q \geq \frac{c}{y-\tau}$ that might choose to become DC users to obtain credit, only households with $q \geq \frac{\alpha\varphi+c}{y-\tau}$ do so, while households with $q \in \left[\frac{c}{y-\tau}, \frac{\alpha\varphi+c}{y-\tau}\right)$ choose cash.

Stage 1. As in the Proof of Proposition 1, we here aim to find the closed form expression for the DC issuer's optimal data access fee to the home lender, $\widehat{\Omega} = m(\theta) E[q|\, dc]\tau$, and use it to show that $\frac{\partial\widehat{\Omega}}{\partial\theta} > 0$, which implies that setting $\theta = 1$ is privately optimal for the DC issuer.

The definitions of the integrals in the expression for the mass of revealed borrowers, depend on whether the $q = 1$ and $\varphi = 2$ household chooses the DC or not. From (38), we have that this household chooses the DC if $y > 2\alpha + c + \tau$ and chooses cash if $y < 2\alpha + c + \tau$. We note from (6) and (5) that $\alpha \in \left[0, \frac{1}{6}c\right]$ and $\tau \in \left[0, \frac{1}{3}c\right]$, which implies that $2\alpha + c + \tau \in \left[c, \frac{5}{3}c\right]$. Given $c < y < \frac{4}{3}c$ in the case of Proposition 2, this means that both $y > 2\alpha + c + \tau$ and $y < 2\alpha + c + \tau$ are possible and should be investigated.

If $y < 2\alpha + c + \tau$, then even at the highest quality ($q = 1$) there are still households who prefer cash over DC. In this case, the mass of revealed borrowers is given by:

$$m(\theta) = \theta \int_{\frac{c}{y-\tau}}^{1} \left(\frac{q(y-\tau) - c}{\alpha}\right) dq = 2\theta\left(1 - \frac{\alpha + c}{y - \tau}\right) \tag{39}$$

and $E[q|\, dc]$ is

$$E[q|\, dc] = \int_{\frac{c}{y-\tau}}^{1} q\left(\frac{q(y-\tau)-c}{\alpha}\right) dq = \frac{(y-c-\tau)^2 (2(y-\tau)+c)}{6\alpha (y-\tau)^2} \tag{40}$$

where we note that the terms $\left(\frac{q(y-\tau)-c}{\alpha}\right)$ derive from setting (38) to equality and solving this cash-DC indifference condition to φ.

Instead, when $y > 2\alpha + c + \tau$ the mass of revealed borrowers is given by:

$$m(\theta) = \theta\left(\int_{\frac{c}{y-\tau}}^{\frac{2\alpha+c}{y-\tau}} \left(\frac{q(y-\tau)-c}{\alpha}\right) dq + \int_{\frac{2\alpha+c}{y-\tau}}^{1} 2dq\right) = \theta\left(\frac{(y-c-\tau)^2}{2\alpha(y-\tau)}\right) \tag{41}$$

and $E[q|\, dc]$ is

$$E[q|\, dc] = \frac{\int_{\frac{c}{y-\tau}}^{\frac{2\alpha+c}{y-\tau}} q\left(\frac{q(y-\tau)-c}{\alpha}\right) dq + \int_{\frac{2\alpha+c}{y-\tau}}^{1} 2qdq}{\int_{\frac{c}{y-\tau}}^{\frac{2\alpha+c}{y-\tau}} \frac{q(y-\tau)-c}{\alpha} dq + \int_{\frac{2\alpha+c}{y-\tau}}^{1} 2dq} = \frac{3(y-\tau)^2 - 4\alpha^2 - 6\alpha c - 3c^2}{6(y-\tau)(y-\alpha-c-\tau)} \tag{42}$$

Taking together (39) and (40) simplifying, and idem for (41) and (42), this means that

$$\widehat{\Omega} = \begin{cases} \theta\tau\frac{\left(3(y-\tau)^2-4\alpha^2-6\alpha c-3c^2\right)}{3(y-\tau)^2} \text{ if } y < 2\alpha+c+\tau \\ \theta\tau\frac{(y-c-\tau)^4(2(y-\tau)+c)}{12\alpha^2(y-\tau)^3} \text{ if } y > 2\alpha+c+\tau \end{cases} \tag{43}$$

Hence, in either case, $\widehat{\Omega}$ is linear and increasing in θ. Thus, $\theta = 1$ is optimal for the DC issuer. ■

Proof of Proposition 3. Per (10), we center attention on the case where $\tau < y - c$, where Proposition 2 implies that high q, low φ households choose to use the DC and, if revealed, receive and accept loan offers, whereas all other households are cash users, remain unrevealed and do not receive loan offers.[73]

Consider $y = \frac{4}{3}c + \varepsilon$ with $\varepsilon \to 0$. In the Proof of Proposition 2, we showed that, even if all borrowers are unrevealed and therefore $E[q|\, u] = E[q] = \frac{3}{4}$, unrevealed borrowers do not take up loans, as there are too many negative NPV borrowers among them and the breakeven loan rate on them is above the return they can earn on their projects. Hence, for $\theta = 0$ (meaning all households are necessarily unrevealed) and $y = \frac{4}{3}c$, the loan market is shut to the unrevealed. Then, for $y = \frac{4}{3}c + \varepsilon$ the same must be true when $\theta > 0$. This is so because $\frac{\partial E[q|u]}{\partial \theta} < 0$ as shown in the Proof of Proposition 1. Indeed, the Proof of Proposition 1, Stage 4, showed that the credit quality of the worst revealed borrower is always better than the average quality of unrevealed borrowers. The same can be shown in the context of Proposition 2; following the same steps at Proof of Proposition 1, Stage 4, but using the expressions in Proof of Proposition 2, the comparison becomes: the credit quality of the

[73]However, the arguments below also apply for $\tau > y - c$, in which case no credit provision takes place in Proposition 2.

lowest quality revealed household necessarily satisfies $q \geq \frac{c}{y-\tau}$, while $\sup E\left[q|\, u\right] = E\left[q\right] = \frac{3}{4}$. Thus, it is sufficient to show that $\frac{c}{y-\tau} > \frac{3}{4}$ and this follows directly from $y \leq \frac{4}{3}c$ and $\tau > 0$. Overall: if the unrevealed do not obtain loans for $\theta = 0$ and $y = \frac{4}{3}c$ then, given $\frac{\partial E[q|u]}{\partial \theta} < 0$, it must be true that for $\theta > 0$ they also do not obtain loans for $y = \frac{4}{3}c + \varepsilon$ with $\varepsilon \to 0$.

The above proves that $y \in \left(\frac{4}{3}c, 2c+\tau\right)$ can contain the outcome observed for Proposition 2. But by extension it also proves that $y \in \left(\frac{4}{3}c, 2c+\tau\right)$ can contain the outcome observed for Proposition 1: for $\theta = 0$, we know that for any $y > \frac{4}{3}c$ it is true that $y > \frac{c}{E[q|u]} = \frac{c}{E[q]} = \frac{c}{3/4}$, and therefore the unrevealed can be made loan offers that satisfy the breakeven condition while leaving positive profit in the hands of a household, if its project is successful. More generally, $\frac{\partial E[q|u]}{\partial \theta} < 0$ implies that whether or not the unrevealed can be made loan offers that satisfy the breakeven condition, depends on θ, with a higher θ raising the breakeven condition for lenders. Moreover, this implies that there can be a threshold value of θ, $\widetilde{\theta}$, above which the outcome of Proposition 2 prevails and below which we get the outcome of Proposition 1.[74]

In cases where there is a $\widetilde{\theta} \in (0,1)$, the outcome will be as in Proposition 2 comes about, since the DC issuer continues to optimally set $\theta = 1$. This follows directly from the optimality of $\theta = 1$ in both Proposition 1 and Proposition 2 in conjunction with the fact that the DC issuer's profit function is continuous: the profit that lenders make from the unrevealed market is 0 at $\theta = \widetilde{\theta}$, because that this the threshold value at which the unrevealed can be made loan offers that satisfy lenders' breakeven condition. When lenders' profit on their loans to the unrevealed market is 0, then so is the DC issuer's profit on that portion of the market at $\theta = \widetilde{\theta}$. That is, there no discontinuous jump in the profit function at $\theta = \widetilde{\theta}$ and for both $\theta < \widetilde{\theta}$ and $\theta > \widetilde{\theta}$ we had already established that DC issuer profit increases as θ rises, so that $\theta = 1$ remains optimal for the DC issuer. ■

Proof of Proposition 4. First, we consider W_h. Since $y_h > 2c + \tau$, the first term in (11) becomes

$$\int_{\frac{1}{2}}^{1} 2\left(qy_h - c\right) dq = \frac{3}{4}y_h - c \tag{44}$$

because all households are borrowers in Proposition 1.[75] Note that this term is strictly positive, given $y_h > 2c + \tau$. The second term in (11) can be found by rewriting the indifference frontier (i.e., (7) holds with equality) to integrate over $\varphi \in [0, 2]$ among DC users,

[74] We note that when θ marginally increases at $\widetilde{\theta}$ (crossing the threshold), the mass of DC users may either increase or decrease. Formally: at $\theta = \widetilde{\theta} - \varepsilon$ with $\varepsilon \to 0$, a household chooses DC over cash if $\alpha\varphi < q\left(\frac{c}{E[q|u]} - \frac{c}{q} - \tau\right)$ (from 18 with R_u given by $\frac{c}{E[q|u]}$ at the threshold while $R_k(q) = \frac{c}{q} + \tau$). Instead, at $\theta = \widetilde{\theta} + \varepsilon$ with $\varepsilon \to 0$, a household chooses DC over cash if $\alpha\varphi < q\left(y - \tau\right) - c$ (from 38). Putting these together, the mass of DC users will be greater for $\theta = \widetilde{\theta} + \varepsilon$ than for $\theta = \widetilde{\theta} - \varepsilon$ if and only if $q\left(\frac{c}{E[q|u]} - \frac{c}{q} - \tau\right) < q\left(y - \tau\right) - c$ which becomes $\frac{c}{E[q|u]} < y$. With $E\left[q|\, u\right] \in \left[\frac{1}{2}, \frac{3}{4}\right]$ and $y \in \left(\frac{4}{3}c, 2c + \tau\right)$ in the intermediate range, this condition can go either way. Intuitively, the threshold pits two forces against each other. On the one hand, when the unrevealed market shuts down, DC use and type differentiation becomes the only path to obtain credit, enticing more households to use DC. On the other hand, cascades unwind, as privacy ceases to be a public good when the threshold is crossed. Since cascades induce more DC use (e.g., Figure 2), their unwinding reduces DC use.

[75] I.e., $f\left(q|\, \text{"borrowing"}\right) = 2$ in (11)

in the same manner that we previously used (7) to integrate over q. Written in this manner $\frac{\alpha\varphi+c}{c}E\left[q|\,u\right]-\frac{1}{2}$ is the probability density function over cash users and $\frac{1}{2}-\left(\frac{\alpha\varphi+c}{c}E\left[q|\,u\right]-\frac{1}{2}\right)$ is the probability density function over DC users.[76] This can be written as $1-\frac{\alpha\varphi+c}{c}E\left[q|\,u\right]$. Hence, the second term in (11) becomes

$$\int_0^2 (\alpha\varphi\theta)\left(1-\frac{\alpha\varphi+c}{c}E\left[q|\,u\right]\right)d\varphi = 2\alpha\theta\left[1-\left(1+\frac{4\alpha}{3c}\right)E\left[q|\,u\right]\right] \tag{45}$$

which we note is always a positive term given $E\left[q|\,u\right]\leq\frac{3}{4}$ and $\alpha<\frac{1}{6}c$, because these imply $\sup\left(1+\frac{4\alpha}{3c}\right)E\left[q|\,u\right]=\frac{11}{12}<1$.

Hence, putting terms together:

$$W_h=\frac{3}{4}y_h-c-2\alpha\theta\left[1-\left(1+\frac{4\alpha}{3c}\right)E\left[q|\,u\right]\right] \tag{46}$$

For W_l, given $y_l\in\left(c,\frac{4}{3}c\right)$ the first term in (11) becomes an integral of $(qy-c)$ over revealed households (because these are the only ones to borrow in Proposition 2). The integrals for revealed households (i.e., θ times DC users) are shown in (39) and (41), and therefore the first term in (11) is[77]

$$\begin{cases} \theta\int_{\frac{c}{y_l-\tau}}^{1}(qy_l-c)\left(\frac{q(y_l-\tau)-c}{\alpha}\right)dq & \text{if } y_l<2\alpha+c+\tau \\ \theta\left(\int_{\frac{c}{y_l-\tau}}^{\frac{2\alpha+c}{y_l-\tau}}(qy_l-c)\left(\frac{q(y_l-\tau)-c}{\alpha}\right)dq+\int_{\frac{2\alpha+c}{y_l-\tau}}^{1}2\left(qy_l-c\right)dq\right) & \text{if } y_l>2\alpha+c+\tau \end{cases} \tag{47}$$

which can be solved to:

$$\begin{cases} \theta\frac{(y_l-c-\tau)^2(2y_l(y_l-c-\tau)+3\tau c)}{6\alpha(y_l-\tau)^2} & \text{if } y_l<2\alpha+c+\tau \\ \theta\left(y_l\left(1+\frac{4\alpha^2}{3(y_l-\tau)^2}\right)+\frac{c^2(y_l-2\tau)}{(y_l-\tau)^2}-2c\right) & \text{if } y_l>2\alpha+c+\tau \end{cases} \tag{48}$$

To come to W_l, we also need an expression for the second term in (11). Here, (39) and (41) represent the integrals for DC users as defined for integration over q, which for the second term in (11) needs to be rewritten for integration over φ, with the probability density

[76] When visualizing the indifference frontier on a plane where φ is on the horizontal and q is on the vertical axis (inverting Figure 2), at every point along φ there is a mass $\frac{\alpha\varphi+c}{c}E\left[q|\,u\right]-\frac{1}{2}$ below the indifference frontier (meaning cash users here; e.g., at $\varphi=0$ households with $q<E\left[q|\,u\right]$ are cash users; take for instance $E\left[q|\,u\right]=\frac{3}{4}$ then the mass of cash users at that point is $\frac{1}{4}$) and since $q\in\left[\frac{1}{2},1\right]$ there is a mass of $\frac{1}{2}-\left(\frac{\alpha\varphi+c}{c}E\left[q|\,u\right]-\frac{1}{2}\right)$ above it, which are DC users (in the earlier example this is $\frac{1}{2}-\left(\frac{3}{4}-\frac{1}{2}\right)=\frac{1}{4}$).

[77] We recall from the Proof of Proposition 2 that the distinction $y_l\lessgtr 2\alpha+c+\tau$ is only relevant when $2\alpha+c+\tau<\frac{4}{3}c$. Given $y_l\in\left(c,\frac{4}{3}c\right)$: if $2\alpha+c+\tau>\frac{4}{3}c$, it follows that $y_l<2\alpha+c+\tau$ for the full range of $y_l\in\left(c,\frac{4}{3}c\right)$ and therefore only the solutions for $y_l<2\alpha+c+\tau$ are applicable.

function taken from setting (38) to equality[78]:

$$\begin{cases} \int_0^{\frac{y_l-\tau-c}{\alpha}} (\alpha\varphi\theta)\left(1-\frac{\alpha\varphi+c}{y_l-\tau}\right)d\varphi & \text{if } y_l < 2\alpha+c+\tau \\ \int_0^2 (\alpha\varphi\theta)\left(1-\frac{\alpha\varphi+c}{y_l-\tau}\right)d\varphi & \text{if } y_l > 2\alpha+c+\tau \end{cases} \tag{49}$$

where the term $\frac{y_l-\tau-c}{\alpha}$ comes from setting $q=1$ and solving to φ (with equality) in (38). These can be solved to

$$\begin{cases} \theta\frac{1}{6\alpha}\left(\frac{(y_l-\tau-c)^3}{y_l-\tau}\right) & \text{if } y_l < 2\alpha+c+\tau \\ \theta\frac{2\alpha}{3}\left(\frac{3y_l-3\tau-4\alpha-3c}{y_l-\tau}\right) & \text{if } y_l > 2\alpha+c+\tau \end{cases} \tag{50}$$

Putting the terms together and simplifying, we can write $W_l = \theta\lambda$ with

$$\lambda = \begin{cases} \frac{1}{6\alpha}\left(\frac{(y_l-\tau-c)^2(y(y-c)+\tau(2c-\tau))}{y_l-\tau}\right) & \text{if } y_l < 2\alpha+c+\tau \\ y_l+\frac{y_l-2\tau}{(y_l-\tau)^2}\left(c^2+2\alpha c+\frac{4}{3}\alpha^2\right)-2(\alpha+c) & \text{if } y_l > 2\alpha+c+\tau \end{cases} \tag{51}$$

where we note that λ is always positive here.[79] ■

Proof of Proposition 5. First, W_h in (13) is necessarily maximized at $\theta = 0$, because then the negative second term (privacy costs) becomes 0.[80] Hence, $\gamma = 1 \Rightarrow \theta^* = 0$. Second, it follows immediately from (14) that W_l is maximized at $\theta = 1$ and therefore $\gamma = 0 \Rightarrow \theta^* = 1$.

The interior solution, $\theta^* \in (0,1)$, is found from $\frac{\partial E[W]}{\partial\theta} = 0$ in (12). For this derivation, it proves useful to define

$$\omega = 6c^2 - 6\frac{c}{\alpha}\frac{3c^2+4\alpha^2}{3c+4\alpha}\left(\alpha - \frac{1-\gamma}{\gamma}\lambda\right) \tag{52}$$

where λ is given by (51). Then, $\frac{\partial E[W]}{\partial\theta} = 0$ yields[81]:

$$\theta^* = \frac{3}{8} - \frac{3}{32}\frac{c^2}{\alpha^2} + \frac{\alpha^2\omega\sqrt{3}\sqrt{(48\alpha^2+3c^4+40\alpha^2c^2)(48\alpha^2c^2+\omega^2)}}{32(48\alpha^2c^2+\omega^2)} \tag{53}$$

[78] To be precise, the probability density function of cash users becomes $\frac{\alpha\varphi+c}{y_l-\tau} - \frac{1}{2}$ and the probability density function of DC users is $\frac{1}{2} - \left(\frac{\alpha\varphi+c}{y_l-\tau} - \frac{1}{2}\right) = 1 - \frac{\alpha\varphi+c}{y_l-\tau}$. Note that for the $y_l < 2\alpha+c+\tau$ case, integration now only goes up to $\varphi = \frac{y_l-\tau-c}{\alpha}$, because beyond that point there are only cash users (and hence no mass of DC users).

[79] In particular, $\frac{(y_l-\tau-c)^2(y(y-c)+\tau(2c-\tau))}{y_l-\tau}$ is immediately positive, while $\inf_{y=2\alpha+c+\tau}\left\{y_l - 2(\alpha+c) + \frac{6c^2+12\alpha c+8\alpha^2}{3(y_l-\tau)} - \frac{y(3c^2+6\alpha c+4\alpha^2)}{3(y_l-\tau)^2}\right\}$ gives $\frac{2\alpha(4\alpha(\alpha+\tau)+c(2\alpha+3\tau))}{(2\alpha+c)^2} > 0$.

[80] This can also be seen by replacing $R_u(\theta)$ from (9) and rewriting the last term in (46) to $-\alpha\theta\left(1 - E[q|u] - \frac{3}{4}\frac{\alpha}{c}E[q|u]\right)$ where it is sufficient to consider this term at the suprema of α and $E[q|u]$ to show that it is always negative overall (i.e., if it would ever be positive, it would have to be at these suprema). These suprema are $\alpha = \frac{1}{6}c$ and $E[q|u] = \frac{3}{4}$ so that the supremum of the term becomes $-\alpha\theta\left(1 - \frac{3}{4} - \frac{3}{4}\frac{1}{6}\frac{3}{4}\right)$ where $1 - \frac{3}{4} - \frac{3}{4}\frac{1}{6}\frac{3}{4} > 0$.

[81] A second, negative root solution can be excluded because it implies $\theta^* < 0$.

which we can further rewrite to

$$\theta^* = \frac{3}{32}\left[4 - \frac{c^2}{\alpha^2} + \frac{\alpha^2\omega}{\sqrt{3}}\sqrt{\frac{48\alpha^2 + 3c^4 + 40\alpha^2c^2}{48\alpha^2c^2 + \omega^2}}\right] \tag{54}$$

■

Proof of Proposition 6. We first show that $\frac{\partial\theta^*}{\partial y_h} = 0$ and $\frac{\partial\theta^*}{\partial y_l} > 0$. Here, $\frac{\partial\theta^*}{\partial y_h} = 0$ can be directly seen from the fact that y_h does not appear in (51), (52), and (54). More generally, in the expression for $E[W]$, y_h only affects W_h, but does so linearly and without an interaction term with θ, as can be seen from (13). Next, $\frac{\partial\theta^*}{\partial y_l} > 0$ is derived by using the $MinValue$ function in *Mathematica* to find the infima of $\frac{\partial\lambda}{\partial y_l}$ over the allowable parameter range ($c \geq 1$, $\alpha \in \left(0, \frac{c}{6}\right]$, $\tau \in (0, y_l - c]$, $y_l \in \left[c, \frac{4}{3}c\right]$). This gives $\inf \frac{\partial\lambda}{\partial y_l} = 0$ where this infimum only occurs in the case where $\tau = 0$, while $\tau > 0$ in our model. It follows that $\frac{\partial\lambda}{\partial y_l} > 0$. This, in turn, implies $\frac{\partial\theta^*}{\partial y_l} > 0$ from (12) and (14). This maps directly to the first sentence of Proposition 6, where $\frac{\partial\theta^*}{\partial\overline{y}} > 0$ and $\frac{\partial\theta^*}{\partial s} < 0$, because θ^* only responds to y_l and not y_h, and a higher $\overline{y}$ (for given s) means a higher y_l while a higher s (for a given $\overline{y}$) means a lower y_l. Instead, when the lower end of the distribution, $\overline{y} - s$, is high enough that the unrevealed market remains open (as per Propositions 1-3), then following Proposition 5, $\theta^* = 0$. ■

B Generalized distribution of y [TO BE DONE]

Instead of binary y in Section 4, now consider a distribution of y which includes the intermediate spectrum (with θ thresholds for opening/closing unrevealed market): **numerical analysis to be done in MatLab**.

Note: for analysis we can use the three expressions of welfare, according to the three ranges of y depicted in Table 1. Replacing y_h and y_l by y in, respectively, (13) and (14), we now have that over the range $y \in \left(c, \frac{4}{3}c\right)$, welfare is given by (14), while for $y > 2c + \tau$ it is given by (13). Over the range $y \in \left(\frac{4}{3}c, 2c + \tau\right)$, welfare is given by

$$\begin{array}{l} (13) \text{ if } \theta < \widetilde{\theta} \\ (14) \text{ if } \theta > \widetilde{\theta} \end{array} \tag{55}$$

where $\widetilde{\theta}$ is found by solving

$$E\left[q|\, u\left(\widetilde{\theta}\right)\right] y = c \tag{56}$$

because when this equation holds with equality, we are exactly at the point where the home lender can expect 0 profit from lending to the unrevealed market at the highest rate that borrowers are willing to accept, y. Here, $E\left[q|\, u\left(\widetilde{\theta}\right)\right]$ comes from replacing θ with $\widetilde{\theta}$ in (30).

Aim: figures showing how socially optimal policy responds to changes in the mean and variance of the distribution.

C Quadratic privacy costs

This extension considers how our model changes when we use a quadratic instead of a linear functional form for privacy costs. We here replace equation (1) with

$$u(q, \varphi) = q \max \{(y - \tau I - R), 0\} - \alpha \varphi \theta^2 \tag{57}$$

where the privacy cost term is quadratic in θ: as the DC becomes more intrusive, household experience a more than proportional increase in their privacy costs. This change does not affect any of the other expressions before Proposition 1. We follow some of the same steps as in Appendix A but, for brevity, do not reproduce the intermediate steps here.[82]

For Proposition 1, nothing changes in Stages 3 - 7 of the game depicted in Section 2.5. In Stage 2, the derived indifference frontier in (7) now becomes

$$\theta \varphi \leq \frac{c}{\alpha} \left(\frac{q}{E[q|u]} - 1 \right) \tag{58}$$

where for $\theta = 1$ we have that (7) and (58) are equivalent. This means that if $\theta = 1$ remains optimal for the DC issuer, then nothing else changes in Proposition 1. Using the same steps as in the proof of Proposition 1, we find that $\theta = 1$ remains optimal for the DC issuer (i.e., the derivative for the closed form expression for optimal data access fees to θ is unambiguously positive).

However, for Proposition 2, outcomes are affected. In particular, as equation (38) now becomes

$$\alpha \varphi \theta < q(y - \tau) - c \tag{59}$$

and following the same intermediate steps as in the Proof of Proposition 2, we now arrive at

$$\mu(\theta) = \begin{cases} 2\left(1 - \frac{c+\alpha\theta}{y-\tau}\right) & \text{if } \theta < \frac{y-\tau-c}{2\alpha} \\ \frac{1}{2\alpha\theta}\left(\frac{(y-c-\tau)^2}{y-\tau}\right) & \text{if } \theta > \frac{y-\tau-c}{2\alpha} \end{cases}, \quad m(\theta) = \begin{cases} 2\theta\left(1 - \frac{c+\alpha\theta}{y-\tau}\right) & \text{if } \theta < \frac{y-\tau-c}{2\alpha} \\ \frac{1}{2\alpha}\left(\frac{(y-c-\tau)^2}{y-\tau}\right) & \text{if } \theta > \frac{y-\tau-c}{2\alpha} \end{cases} \tag{60}$$

where we note that for both $\mu(\theta)$ and $m(\theta)$ the terms in the two expressions are equivalent when $\theta = \frac{y-\tau-c}{2\alpha}$. This means that the masses of DC users and revealed borrowers do not portray discrete jumps over θ. However, those masses do have a kink in their response to θ. This implies that this functional form is less well suited to analyze DC design.

But although this functional form lends itself less well to the extended analysis of the baseline, the fact that with quadratic privacy costs $\mu(\theta)$ and $m(\theta)$ respond differently to θ than in the baseline speaks to some of the mechanisms underlying the model. In particular, in the low y case (Proposition 2), we now have that $\frac{\partial \mu(\theta)}{\partial \theta} < 0$.[83] In this case, the quadratic cost makes marginal privacy costs (from higher θ) large enough that more households are repelled from DC use in response to increasing privacy costs than are attracted into the DC due to the higher loan rates on the unrevealed. This does not imply that the mass of revealed borrowers also decreases as θ increases, however. When $\theta > \frac{y-\tau-c}{2\alpha}$, we here

[82] Calculations are available in a *Mathematica* file that is available on request.

[83] This is true for both expressions of $\mu(\theta)$, i.e., regardless of $\theta \lessgtr \frac{y-\tau-c}{2\alpha}$.

have that $\frac{\partial m(\theta)}{\partial \theta} = 0$, as the decline in the mass of DC users is exactly offset by the higher probability of revelation per DC user, so that the mass of revealed borrowers remains the same. When $\theta < \frac{y-\tau-c}{2\alpha}$, we have $\frac{\partial m(\theta)}{\partial \theta} > 0$ like in the baseline, because the decline in the mass of DC users more than offset by the higher probability of revelation per DC use.[84]

Interestingly, when privacy costs are quadratic, $\theta = 1$ is no longer privately necessarily optimal for the DC issuer in the Proposition 1 setting. Using the same setps as for Proposition 1 we obtain closed form solutions for optimal data access fees

$$\widehat{\Omega} = \begin{cases} \frac{1}{\theta}\frac{\tau(2(y-\tau)+c)(y-\tau-c)^4}{12\alpha^2(y-\tau)^3} \text{ if } y < 2\alpha\theta + c + \tau \\ \frac{\tau}{3(y-\tau)^2}\left(\left(3(y-\tau)^2 - 3c^2\right)\theta - 6\alpha c\theta^2 - 4\alpha^2\theta^3\right) \text{ if } y > 2\alpha\theta + c + \tau \end{cases} \tag{61}$$

where it is important to note that for $\theta \to 0$, we always have $y > 2\alpha\theta + c + \tau$, given (10). This means that there is no singularity at $\theta \to 0$, which seems possible from the first entry in (61), because the second entry in (61) always applies for $\theta \to 0$.

From (61), we can derive that both $\theta = 1$ and $\theta \in (0, 1)$ are possible as privately optimal for the DC issuer, depending on parameter values. For example, when $\alpha \to 0$, we have that the second entry in (61) always applies (as $y > c + \tau$ (10)) and becomes $\frac{\tau}{3(y-\tau)^2}\left(\left(3(y-\tau)^2 - 3c^2\right)\theta\right)$ which monotonically increases as θ rises and therefore $\theta = 1$ is optimal for the DC issuer. Instead, when parameters satisy $y < 2\alpha + c + \tau$, then for $\theta \to 1$, the first entry in (61) applies, where $\widehat{\Omega}$ monotonically decreases as θ rises and therefore $\theta = 1$ cannot be optimal for the DC issuer. As we had already established that $\theta = 0$ cannot be optimal either (as the second entry in (61) always applies when $\theta \to 0$), it follows that $\theta \in (0, 1)$ is privately optimal for the DC issuer.[85]

Intuitively, when the marginal cost of privacy rises as θ increases (because privacy costs are a quadratic function of θ), there can be cases when the increasing θ causes enough exit from DC use that the DC issuer no longer wants to offer a fully revealing DC.

Notably, however, the private optimum certainly remains $\theta = 1$ precisely when the social optimum is $\theta^* = 0$, namely in the good (high y) state described by Proposition 1.[86] When credit access is ubiquitous, the divergence between social and privately optimal policy is extreme.

Instead, when credit access is constrained to revealed households, the comparison between socially and privately optimal policy becomes a nuaned affair. We rederive W_l following the same steps as in the proof of Proposition 4. The socially optimal policy can now be either $\theta^* = 1$ or an interior $\theta^* \in (0, 1)$, depending on parameter values.[87] [Possibly to be added: a numerical check of whether parameterizations now exist where θ^* is higher than private optimum. If yes, explore what might be the intuition for this].

[84] As the infimum of $\frac{\partial m(\theta)}{\partial \theta}$ then occurs at $\theta = \frac{y-\tau-c}{2\alpha}$ where $\frac{\partial m(\theta)}{\partial \theta} = 0$ (i.e., $\frac{\partial m(\theta)}{\partial \theta}$ for any $\theta < \frac{y-\tau-c}{2\alpha}$).

[85] To be precise, privately optimal θ will be either one of the two solutions to FOC of the second entry in $\theta \in (0, 1)$ or will be at the point where $y = 2\alpha\theta + c + \tau$, which is $\theta = \frac{y-\tau-c}{2\alpha}$. For a given parameterization, numerical investigation would indicate which of these solutions gives the highest $\widehat{\Omega}$, which would be the private optimum.

[86] We rederive W_h and find that it continues to be maximized at $\theta^* = 0$, like in the baseline model.

[87] Derivations are available on request from the authors.

D Alternative sources of credit quality data

The baseline model centers on a simple, stark choice between DC and cash as, respectively, partially and completely unrevealing payment technologies. What if instead lenders had access to alternative sources of credit quality data, which need not be purchased from the DC issuer? We consider two possibilities: pre-existing public information and public information that is gathered through an alternative payment system technology.

The first possibility is that there is pre-existing public information about credit quality, such as through credit registries or FICO scores. We assume that this information is freely available to lenders and we can incorporate this into our framework through a revelation draw that occurs before the rest of the game outlined in Section 2.5. That is, before the game begins, some households become revealed and their credit quality is public information. Per our derivations in Appendix A, it follows directly that such households always receive loan offers $R_k(q)$ as in (8). Nothing else changes in the game. Because this additional revelation occurs before household decisions, it has no impact on household payment choice. That is, for the set of households that is not pre-revealed, the baseline model is exactly the same as before. Lender and DC issuer total profits will be affected, but only because the pool of households that is not pre-revealed is smaller. With a renormalization, total profits would remain the same as in the baseline too: i.e., if the probability of pre-revelation is $\xi \in (0,1)$, then the renormalization sets the total mass of households per home lender at $\frac{1}{\xi}$.

The second possibility is that, instead of cash, households have access to another alternative to payment technology that creates public information about credit quality. Household payment choice centers on DC in comparison to this alternative means of payment.[88] This could be deposits or CBDC, for example.[89] Let the probability of household credit quality revelation be θ_{alt} when using this alternative, as compared to θ_{dc} when using the DC. We center attention on the case where $\theta_{alt} < 1$ and we restrict $\theta_{dc} \geq \theta_{alt}$. That is, the alternative will leave some households unrevealed and we focus on a setting where the DC is used for additional disclosure.

Here, (17) becomes

$$q\left(y - \tau I - \left(\theta_{dc} R_k(q) + (1-\theta_{dc}) R_u\right)\right) - \alpha\varphi\theta_{dc} > q\left(y - \tau I - \left(\theta_{alt} R_k(q) + (1-\theta_{alt}) R_u\right)\right) - \alpha\varphi\theta_{alt}$$

which can be written to

$$\alpha\varphi\left(\theta_{dc} - \theta_{alt}\right) < q\left(\left(\theta_{dc} - \theta_{alt}\right) R_u - \left(\theta_{dc} - \theta_{alt}\right) R_k(q)\right) \tag{62}$$

and

$$\alpha\varphi < q\left(R_u - R_k(q)\right) \tag{63}$$

which is identical to (18). That is, the condition that previously identified household payment choice between DC and cash, now identifies household payment choice between DC and the

[88] Competition between three payment technologies (cash, DC, and another alternative) lies beyond the scope of this paper (but see Agur et al. (2022) for an analysis that considers this).

[89] Note, however, that we restrict attention to *public* information gathered through such payment systems. Bank deposits, in particular, could also create lender-specific data on a subset of households (namely, the lender's own depositors). This would significantly complicate the model and lies outside the scope of the current analysis.

alternative technology.

Similarly, following through on other steps in Appendix A, nothing changes about the DC issuer optimization. For instance, (33) sees θ replaced with $\theta_{dc} - \theta_{alt}$ and therefore $\theta_{dc} = 1$ remains privately optimal.[90] Hence, this alternative payment technology does not add novel insights to the model: its outcomes are unchanged compared to the baseline.[91]

[90] I.e., θ_{alt} becomes an added constant in the optimization to θ_{dc}, which does not affect the optimality of $\theta_{dc} = 1$.

[91] Welfare analysis is also the same, except that socially optimal policy, θ^*_{dc}, faces a lower bound constraint at θ_{alt}.

E Alternative data selling strategies [TO BE DONE]

The structure that we have considered for the DC issuer, where it sells the data non-exclusively to lenders, is privately optimal to the DC issuer. Consider structures whereby either: a) the DC issuer sets up its own lender (assume zero setup cost and lender is otherwise equivalent to other lenders) and provides data for free to this lender but not to any other lender; or b) the DC issuer sells the data exclusively to the highest bidding lender. These structures "unravel" in the sense of the first sentence of Proposition 2: a "monopolistic" lender with a time inconsistency problem. The situation is a little different than in the high τ case of Proposition 2, in that there is no pure monopoly here, because borrowers can still turn to uninformed lenders and obtain the unrevealed rate. However, given the unique ownership of the data by either the DC's subsidiary lender (case a) or the lender that obtains exclusive access (case b), at Stage 5 of the game, this lender will charge all revealed borrowers $\varepsilon \to 0$ below the unrevealed rate. The revealed borrowers are "cornered": they cannot do better than take up this offer and this means that effectively all value from becoming revealed vanishes to them. Given this outcome, at Stage 2 all households choose cash, because there is only a privacy cost to the DC but no prospective loan rate gain. Lenders' inability to precommit to lower loan rates (including the DC subsidiary or the lender to which it sells exclusively) means that lenders cannot entice any household to use the DC. Hence, the DC issuer makes zero profit in both cases a and b and therefore is better off with the structure considered in the baseline, with non-exclusive access.[92,93]

[92] Add: also note that block selling is optimal. Household-by-household sale of data opens the door to information gaming by lenders. Simplified example: consider a setting with X households, whose average quality is known. If the home lender purchases data on $X - 1$ households, it can always infer the credit quality of the Xth household.

[93] Could generalize this appendix: appendix on why the way we set up the problem for the DC issuer - with non-exclusive block sale of data - is optimal. Follows directly from: cannot get more profit than full profit on revealed borrowers, which DC issuer manages to extract in equilibrium.