Bank Competition and Household Privacy in a Digital Payment Monopoly*

Itai Agur — IMF, iagur@imf.org

Anil Ari — IMF, aari@imf.org

Giovanni Dell'Ariccia — IMF, gdellariccia@imf.org

December 14, 2023

Abstract

Lenders can exploit households' payment data to infer their creditworthiness. When households value privacy, they then face a tradeoff between protecting such privacy and attaining better credit conditions. We study how introducing an informationally more intrusive digital payment vehicle affects households' cash use, credit access, and welfare. A tech monopolist controls the intrusiveness of the new payment method and manipulates information asymmetries among households and oligopolistic banks to extract data contracts that are more lucrative than lending on its own. The laissez-faire equilibrium entails a digital payment vehicle that is more intrusive than socially optimal, providing a rationale for regulation.

Keywords: Privacy; Financial intermediation; BigTech; Data regulation.

JEL codes: D82, E41, G21, G28

*We would like to thank Yuteng Cheng, Florian Heider, Peter Hoffmann, Michael Junho Lee, Sergio Vicente, and Brandon Tan for discussions and audiences at the 2023 Swiss Winter Conference on Financial Intermediation, the 2023 Luiss-CEPR Finance Workshop, the 2023 Oxford Financial Intermediation Theory Conference, the CEPR-Bocconi Conference on The Future of Payments and Digital Assets, the IMF's 5th Annual Macro-Financial Conference, the BIS, Collegio Carlo Alberto Turin, 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

New payment technologies are making it increasingly easy to collect and store information about consumer behavior, while network externalities limit the number of viable payment providers and endow them with significant market power in the collection of personal data.[1] This data can be parsed to infer a variety of personal characteristics that corporations can profitably exploit, including by deriving credit ratings based on an individual's spending behavior. In credit markets, such use of payment data can reduce informational asymmetries and unlock market segments where adverse selection may otherwise hinder the provision of credit.[2] But this may come at the expense of personal privacy, an acute concern for many households.[3] Further, market power in data collection may give payment providers an edge that alters the competitive landscape in banking markets.

How will markets navigate tradeoffs between credit provision and privacy? Who will be the winners and losers of the introduction of new forms of data collection? Does the rise of these new technologies call for government intervention? In this paper, we shed light on these questions by building a model in which a digital payment issuer competes against an anonymous form of payment, cash. This presents households with a choice between one payment method that precludes information collection and another with an adjustable informational

[1]Payment data is increasingly migrating to BigTechs. E.g., when using a bank card at a store, the location and type of store are visible to the bank, whereas for Amazon purchases with such a card "Amazon.com" appears on the charge and only Amazon observes the purchase details. Moreover, several BigTechs encourage payment data migration beyond their platforms. E.g., Amazon rewards consumers that share receipts of purchases outside Amazon; Alibaba and WeChat provide widely used means of payment in China; and Meta has explored digital currency issuance (Libra and Diem).

[2]Indeed, there is growing evidence that new means of payment, including those created by BigTechs, expand credit access. See, e.g, Agarwal and Assenova (2022), Allen et al. (2021), Babina et al. (2022), Beck et al. (2022), Berg et al. (2020), Bian et al. (2023b), Dalton et al. (2023), D'Andrea and Limodio (2023), Doerr et al. (2023a), Frost et al. (2020), Ghosh et al. (2022), Hau et al. (2019), Huang et al. (2020), Ouyang (2022), and Sarkisyan (2023).

[3]Although the value of digital privacy in general is empirically debated (Acquisti et al., 2016; Athey et al., 2017; Bian et al., 2023a; Chen et al., 2021; Goldfarb and Que, 2023; Tang, 2023), for payments and lending specifically, most studies find an important role for privacy. See, e.g., Bijlsma et al. (2022, 2023), Borgonovo et al. (2021), Brits and Jonker (2023), Choi et al. (2023), Cunliffe (2023), ECB (2021), Engels et al. (2022), Hu et al. (2023), and Li (2023) on privacy and payment choice; and Tang (2023) and Doerr et al. (2023b) on borrowers' willingness to pay to limit the intrusiveness of data disclosure in loan applications. See also Berg et al. (2022) (pp.197-198) for a discussion of the empirical literature on privacy preferences and lending outcomes.

intrusiveness: the issuer of the digital payment vehicle can decide what consumer-spending data to collect and retain.[4] For instance, for each transaction the only data collected and retained could be the vendor and the amount spent, similar to what happens with credit cards, or instead reach the level of the specific items bought. Further, the data collector could choose what segments of payment data to connect to other sources of information about a specific individual: e.g., payment data could be linked to users' activity on social media accounts or ecommerce platforms.[5] Alternatively, the digital payment provider could introduce a digital currency with customizable privacy features.[6] In the model, intrusiveness is a choice variable for the digital payment provider that determines the probability with which the digital payment vehicle reveals the creditworthiness of its users.

Households are consumer-entrepreneurs that differ in their creditworthiness and their care for privacy, both of which are private information.[7] They are endowed with a project but need external financing in order to realize it. The digital payment provider sells the information it collects about consumer behavior to banks, which can exploit this data to reduce adverse selection when competing for loans to borrowers in an imperfectly competitive market. A stylized spatial model provides individual banks with market power (we assume that each bank has a cost advantage in lending to its "home" market).

When optimizing households choose between using cash and the digital payment vehicle, they take into account how being observed by the digital payment issuer affects their personal cost in terms of loss of privacy and the potential revelation of their creditworthiness (and thereby their borrowing conditions). Similarly, the digital payment provider takes this optimizing behavior into account when determining the intrusiveness of the payment vehicle.

[4] Our analysis does not hinge on cash being the alternative or payments being the only source of credit quality data. As shown in Appendix B.1, the key assumption is that the issuer's means of payment reveals *more* credit quality information, not that it is the unique source of such information.

[5] As has been the case for, respectively, WeChat Pay and Alipay in China

[6] Depending on their design, digital currencies can combine attributes of deposits and cash, including the extent of privacy (Agur et al., 2022).

[7] Studies reporting extensive heterogeneity in digital privacy preferences include Bian et al. (2023a), Budak et al. (2023), Collis et al. (2022), Goldfarb and Tucker (2012), Jiang et al. (2023), Lin (2022), Lin and Strulov-Shlain (2023) and Prince and Wallsten (2022).

How much data the provider manages to collect depends on the intrusiveness of the vehicle and on the portion of consumers that chooses to use it.

The model's equilibrium delivers four main results. First, the digital payment provider makes the payment vehicle as intrusive as possible (leading to certain revelation of creditworthiness type for all users). The payment provider's optimal strategy stems from adverse selection, which it exploits to instigate a partially unravelling Lemons' Market. This plays on the way that households sort into means of payment. Households with low privacy costs and high creditworthiness want to use the digital payment vehicle, while households with high privacy costs and low creditworthiness opt for cash. In between, there are households that would choose cash if everyone chose cash, but are swept along in a disclosure cascade, because high quality households' departure from cash worsens the average creditworthiness of the pool of borrowers using cash and hence raises their equilibrium loan rate.[8] As a result of households' disclosure externality, the market share of the digital payment vehicle increases with intrusiveness, despite the higher privacy cost experienced by individual users.[9]

A second result is that the digital payment provider maximizes fee revenue by offering free data access to banks outside their home market. Banks compete à la Bertrand in the credit market and (under symmetric information) a price-limit equilibrium emerges wherein each bank serves and extracts surplus from borrowers in its home market and unsuccessfully contests other markets. Hence, the home bank stands to lose the most (and thus is willing to pay the most) from being cut out of private information about borrowers' creditworthiness, and its willingness to pay for data is maximized by the threat of an asymmetric equilibrium when out-of-home-market information is free to all banks.

The third result is that the digital payment provider prefers to sell information to the banks rather than becoming a lender and walling off the information for its own use. This stems from the fact that the presence of cash makes its monopoly power partly contestable. Should the provider become a lender itself, it would use the payment information

[8] Despite these strategic complementarities, there is a unique equilibrium.

[9] Appendixes B.2 and B.3 investigate the robustness of these model dynamics to functional form choices.

to gain monopoly power over loan markets and thus extract all the surplus from borrowers/consumers. Foreseeing this, consumers would all use cash, denying the digital provider any informational advantage versus competing lenders.[10] Competition among banks here acts as a commitment device for the payment provider to share some of the surplus with borrowers. The ability to credibly commit to lower loan rates would assuage this outcome, but in the absence of perfect commitment devices, this provides a counterbalance to the integration of data collection and credit provision in a data monopoly.[11]

These results come about in an economy where enough borrowers have positive NPV projects that in equilibrium every project is financed. Borrowers with negative NPV projects that choose to transact with cash get financed because, even after disclosure cascades, there remain enough borrowers with high creditworthiness and high care for privacy that also choose cash. This in turn implies that, while there will be winners (households with a low care for privacy and high creditworthiness) and losers (other households), the aggregate effects of introducing the digital payment vehicle are negative: relative to a cash-only world, aggregate output is unchanged and there are positive privacy costs. It follows that under these circumstances a social planner would always prefer a non-informative form of payment.

To speak to the potential benefits of an information producing payment vehicle, we consider an economy where under certain circumstances information disclosure leads to greater aggregate credit access. Put simply, we consider states of the world under which the absence of information about individual borrowers leads to a market shutdown. When the loan market for undifferentiated households freezes, type differentiation becomes of social value.[12] A more intrusive data collection now gives households with relatively good credit quality a way to stand out and obtain credit. This raises total credit provision and aggregate welfare.

We use this framework to draw insights about the optimal regulation of a data monopoly, which constitutes our fourth key result. Regulation is envisaged as a slow-moving policy,

[10] A similar reasoning applies if the payment provider sells its data exclusively to one bank.

[11] As opposed to, e.g., economies of scope favoring integration (Huang, 2022; Rishabh and Schäublin, 2021).

[12] We also investigate intermediate states in which payment data collection determines whether lending to undifferentiated households is sustained: a sufficiently intrusive design collapses that market.

determined before the economic state materializes, and implemented as a constraint on the intrusiveness of the monopolist's data collection. Optimal regulation here trades off the costs of intrusiveness when credit provision is assured against its benefits during a credit crunch.

A key insight is that regulation is most needed when the competing means of payment both have sizable market shares. This is because balanced market shares arise from intermediate privacy preferences, under which disclosure cascades are both large and painful enough to matter. We also analyze how optimal regulation responds to market power in lending and to the distribution of returns across states.

Furthermore, we find that there can be a role for a second policy instrument. Optimal policy targets the amount of revealed household data, which consists of an extensive (the share of households opting for disclosure) and an intensive margin (the extent of revelation about such households). Regulation affects both of these margins. But because there are two margins, the policy maker can attain higher welfare if a second instrument is available. We explore the example of a tax or subsidy on households that opt for the monopolist's payment method, which impacts only the extensive margin of disclosure. The reason that a subsidy can potentially be socially optimal is that there is underprovision of household data in the bad state, even with a maximally intrusive digital payment system design: bank market power implies markups that repel some positive NPV households from disclosure, leaving their projects unfinanced in this state and reducing aggregate welfare. Depending on parameter values, either a tax or a subsidy can be optimally combined with regulation.[13]

We highlight the extent to which the payment data monopoly amplifies externalities in an extension wherein the digital payment system is owned by a consortium of banks rather than a monopolist.[14] The welfare outcome is in between the monopolist and the social planner: the bank consortium neither actively exploits nor leans against the disclosure externalities. A similar outcome is seen in another extension wherein households own their data, which

[13]Relatedly, the two instruments can optimally either be used as substitutes (e.g., a tax and looser regulation than if regulation acts alone) or complements (e.g., a subsidy and looser regulation).

[14]The consortium setting is inspired by Allen and Gale (1999).

relates to the data porting and open banking policies observed in an increasing number of countries (OECD, 2023).

The remainder of the paper is organized as follows. The next section reviews the related literature. Section 3 presents the setup of the model. Section 4 derives the equilibria wherein the data monopolist sells the data to lenders and Section 5 shows why it chooses to do so instead of engaging in credit provision itself. Section 6 analyzes welfare and socially optimal policy design. Section 7 extends to lender ownership of the data monopolist, to help distill the monopolist's role in the baseline model. Proofs can be found in Appendix A and additional model extensions in Appendix B.[15]

2 Related Literature

Our paper relates to a growing literature on emerging financial technologies in payments and credit provision.[16] Like us, Parlour et al. (2022b) and He et al. (2023) consider borrowers that are heterogeneous in both credit quality and a second dimension (respectively, bank affinity and privacy), and data externalities among such borrowers take center stage.[17] In Parlour et al. (2022b), FinTech payment providers compete against a bank and externalities stem from coupling data garnered from payment services with credit provision.[18] He et al. (2023) also consider lending market competition between banks and FinTech entrants, and show that household data ownership is suboptimal as individual data sharing decisions lead

[15] These extensions are: alternative sources of credit quality data (B.1); correlated dimensions of household heterogeneity (B.2); an alternative functional form for privacy costs (B.3); fees / subsidies by the monopolist on the use of its payment vehicle (B.4); the interaction between data collection incentives and monetary policy (B.5); and household data ownership (B.6). Lastly, Appendices B.7 and B.8, respectively, contain additional derivations for Sections 4 and 6.

[16] Much of this literature builds on the classical literature on optimal disclosure (Akerlof, 1970; Grossman and Hart, 1980; Jovanovic, 1982; Milgrom, 1981).

[17] These setups aligns with empirical findings that additional dimensions of agent heterogeneity (in addition to hidden quality) avoid a full "unraveling" equilibrium where all agents opt to disclose private information (Bond and Zeng, 2022; Jin et al., 2021; Jin and Vasserman, 2021; Soleymanian et al., 2021).

[18] On asymmetric information and data externalities, see also Acemoglu et al. (2022), Bergemann and Bonatti (2023), Bergemann et al. (2022), Choi et al. (2019), Cicala et al. (2022), Garratt et al. (2021), Jones and Tonetti (2020) and Parlour et al. (2022a).

to aggregate credit quality inferences by lenders.[19]

Different from these studies, we consider a data monopolist that stands "upstream" and controls the flow of information, actively exploiting information asymmetries and data externalities. In essence, we model a BigTech, whereas Parlour et al. (2022b) and He et al. (2023) model FinTechs.[20] Unlike the FinTechs, our BigTech determines the extent of intrusiveness of its data collection, while it competes with cash, which allows households to opt out of its data web. This lays bare the incentives of a data monopolist to play off households and banks against each other, including by staying above the fray of the credit market and selectively offering free data access to some banks.

Several papers consider the role of a Central Bank Digital Currency (CBDC) in the preservation of privacy. In Garratt and Van Oordt (2021), the social value of CBDC derives from the fact that privacy is a public good: taking actions to protect digital privacy is costly to a consumer, while firms use data collected through payments to price discriminate future consumers.[21] In Garratt and Lee (2021), the CBDC 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. 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 counteracting payment

[19] 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.

[20] Information flow on ecommerce platforms provides a different angle on BigTechs. In Markovich and Yehezkel (2023) privacy cost heterogeneity underlies the suboptimality of giving households 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. Bouvard et al. (2022) focus on monitoring efficiency differences between an ecommerce platform and banks in the competition for loans towards merchants that are active on the platform. In Gambacorta et al. (2022) platforms have an enforcement advantage over banks, because defaulting firms can face exclusion from the platform. On ecommerce data collection incentives, see also Charlson (2021), Choe et al. (2023), Cong and Mayer (2023), Fainmesser et al. (2023), Gambato and Peitz (2023), Ichihashi (2023), Liu et al. (2023b), Kim (2021) and Petropoulos et al. (2023).

[21] Privacy in payments can also be a public good when there are no data externalities, such as when there is the risk of identity theft (Kahn et al., 2005; Kahn and Roberds, 2008).

data monopolies.[22] Different from us, these papers do not focus on information asymmetries in credit markets.[23]

3 Model

We consider an economy populated by three sets of agents. Consumer-entrepreneurs, to which for simplicity we refer as "households", form the first set. Second, there is a monopolist digital currency issuer.[24] Lenders comprise the third set of agents.[25] All agents are risk neutral. The economy is distributed across $N \geq 3$ connected islands. Each island hosts one lender and a continuum of households with mass 1. A lender on the same island as a borrower is called the borrower's "home lender", and borrowers on the same island as a lender are referred to as the lender's "on island" borrowers. All islands share the same physical currency (cash) and digital currency issuer.

3.1 Households

Each household is characterized by three variables: the household's creditworthiness and preference for privacy, which are private information, and the island where the household lives, which is publicly observable. The last is self-explanatory. Let us turn to the first two. Each household is born with an investment project that can only be brought to fruition with financing from a lender. All projects yield the same payoff, y, when successful and 0 otherwise. However, the probability that a project succeeds, q, differs across households. We

[22] On new forms of "smart money", see also Kahn and van Oordt (2023), Schneider and Taudien (2023) and Tan (2023).

[23] Brunnermeier and Payne (2022, 2023) do incorporate credit markets, but focus on enforcement and search costs rather than disclosure of hidden qualities.

[24] The model's setup and results apply to digital payment monopolists broadly but, to fix ideas, we call the monopolist that stands at the core of our model a "digital currency issuer". Its business model, based on the collection and sale of data, is akin to Meta Platforms, while the notion of a digital currency resembles the Libra and Diem initiatives. We do not model what gives rise to a monopoly here, such as network effects in means of payment (Agur et al., 2022). On the endogenous choice of a data monopolist to launch a digital currency, see also Chiu and Wong (2022) and Guennewig (2022).

[25] We here use the term "lenders" instead of banks, because the way they are modeled (Section 3.3) is general enough to encompass nonbank lenders, including FinTechs.

refer to q as a household's credit quality or creditworthiness. Households also differ in their preference for privacy or, put differently, their perceived cost in having their consumption patterns scrutinized. Formally, each household attaches disutility φ to its loss of privacy. A household's creditworthiness and privacy preferences are independent of each other.[26] On each island, the mass 1 of households is uniformly distributed on a two-dimensional plane with $q \in \left[\frac{1}{2}, 1\right]$ and $\varphi \in [0, 2]$.[27]

3.1.1 Household consumption

Households are born with an endowment that they use towards consumption. To avoid complicating the model excessively, we assume that this consumption is needed to meet subsistence needs (i.e., none of the endowment can be saved or used as collateral when requesting credit). However, an individual household's consumption pattern, if observed, may carry information about her creditworthiness.

Each household chooses between cash and the digital currency (DC) to pay for consumption.[28] Cash fully protects the household's privacy. If instead the household uses the DC then with some probability, θ, the DC issuer is able to infer the household's creditworthiness.[29] The household experiences privacy disutility φ from being observed by the DC issuer.[30] Households understand the *extent* to which their behavior is observed by the DC issuer. That extent, which is captured by θ, determines both the household's privacy disu-

[26] Appendix B.2 considers an extension where q and φ are (either positively or negatively) correlated.

[27] We choose $q \in \left[\frac{1}{2}, 1\right]$ instead of $q \in [0, 1]$, as the latter creates additional complexities (individual breakeven loan rates go to infinity as $q \to 0$, which complicates the identification of simple parameter conditions to characterize equilibria; see, e.g., Table 1 and footnote 90). Using $q \in \left[\frac{1}{2}, 1\right]$ is without loss of generality, because all possibilities on the functioning of the loan market are covered by our framework (Section 4). Given $q \in \left[\frac{1}{2}, 1\right]$, we let $\varphi \in [0, 2]$ instead of $\varphi \in [0, 1]$ to obtain a unit mass of households on each island.

[28] In our setting, it would not make a difference to allow households to use multiple means of payment for consumption: in equilibrium, each household will choose only one, as will be seen from (7). Meyer and Teppa (2023) report that 60 percent of Euro Area consumers do actually conduct all their purchases with only one preferred means of payment.

[29] The model generalizes to partial revelation from other sources than DC use, as discussed in Appendix B.1, as long as DC use entails additional potential revelation.

[30] For the development of a continuous privacy metric and its empirical application, see Dekel et al. (2023).

tility and the probability of type revelation.[31] But, to avoid having to deal with strategic consumption behavior, we assume that households are unable to predict how their consumption choices affect their credit score (put differently, households have no visibility into the algorithm used to extract creditworthiness information from consumption patterns). For a practical example, consider the case of a social media company that issues a DC and that can choose whether to link households' payment and social media data.[32] The decision to link these data is observed by households, and disliked (to varying degrees) from a privacy angle by those who choose to use DC. But the way in which social media chats interact with payment data is likely poorly understood by households.[33]

Note that because of this setup the only meaningful decision households make at the consumption stage is about the form of payment they use.

3.1.2 Household borrowing and payoff

Households are protected by limited liability. They may borrow from any lender. However, borrowing from a lender not located on their home island reduces the payoff of a successful project by $\tau > 0$.[34] This allows us to introduce a degree of lender market power in a simple and tractable form. The benefit in borrowing from the home lender can be interpreted as spatial differentiation (Degryse and Ongena, 2005) or sectoral specialization of lenders that are matched to household projects in the same industry (Allen et al., 2011).

[31] To give a near-future example that helps visualize the tradeoff, consider AI-enabled cameras that watch consumers at in-person BigTech venues (some Amazon-owned supermarkets are already equipped with arrays of cameras, although aimed at facilitating facial recognition shopping). These may, e.g., capture pupil dilation when looking at healthy or unhealthy products. More cameras, set up closer to people's faces, improve the ability of the system to gauge involuntary reactions, which can relate to credit risk. But the more cameras are placed and the closer they are to faces, the more intrusive the system will feel to consumers.

[32] E.g., WeChat has derived credit ratings from linked payment and social media data. DC design is further discussed in Section 3.2.

[33] E.g., social media chats can reveal the motivation for purchases (e.g., whether a liquor purchase is for a party rather than own consumption). However, the household does not consider this when engaging on social media (i.e., the household is assumed not to create a fake chat, like on planning a party, to affect its credit ratings).

[34] To guarantee that limit pricing can be sustained as a Nash equilibrium in Section 4, we additionally assume a breakeven preference in favor of the home lender. An alternative is the approach of Blume (2003), which allows for limit pricing to be supported as a mixed-strategy Nash equilibrium. But we focus on pure strategy solutions here and therefore work with the breakeven preference.

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 the household's project under limited liability, and $-\alpha\varphi\Theta$ is privacy costs. In $q \max\{(y - \tau I - R), 0\}$, I is an indicator variable that equals 0 when a household borrows from its home lender and 1 otherwise; R is the gross loan interest rate charged; and the max operator reflects the assumed limited liability protection.[35] In $-\alpha\varphi\Theta$, $\alpha > 0$ is the relative weight on privacy preferences and Θ represents the probability that the household's credit quality will become revealed: $\Theta = \theta$ if the household chooses DC and 0 if it chooses cash.[36]

For simplicity, the baseline model does not incorporate an additional term to represent an ease-of-transactions benefit of DC relative to cash. Appendix B.8 recalculates the main expressions when including such a term.[37] We reiterate, however, that cash is merely one example of an alternative payment instrument. For example, our modeled alternative to the DC could equally well be a privacy-preserving crypto asset, which may not have a transaction disadvantage relative to the DC.[38]

3.2 Digital currency issuer

The DC issuer is in control of the design of the digital currency. Formally, it chooses the probability, θ, with which it will be able to learn the creditworthiness of individual households

[35] The expression is of the form $q \max\{., 0\}$ because even a successful project could, in principle, violate the limited liability constraint when R is large enough. This will not occur in equilibrium, however, and the max operator therefore becomes moot by optimal play in Section 4.

[36] In the baseline model, privacy costs are therefore linear in the probability of revelation. Appendix B.3 considers quadratic privacy costs.

[37] Appendix B.8 centers on a government tax or subsidy on DC use. A simple (linear) transaction benefit for DC is equivalent to an unfunded subsidy in that setting (i.e., $T < 0$ and $transfers = 0$ in equation 50). On transaction benefits with network effects, see Agur et al. (2022).

[38] On alternative means of payment, see also Appendix B.1.

using the DC.[39] The DC issuer is a monopolist in the provision of digital payments and hence also sets the price at which its collected digital payment data are made available to lenders. Note that in this section we focus on the case of a DC issuer that puts these data up for sale to lenders and does itself not engage in lending. Section 5 considers the possibility that the DC issuer would instead provide credit, either directly or through a subsidiary lender, and shows that this will never emerge as an equilibrium.

The DC issuer may charge lenders differentiated fees for information on borrowers on and off their islands. Here, Ω_{ij} denotes the data access fee that the digital currency provider charges lender i for access to data on households on island j.[40] We define $D_{ij}(\theta, \Omega_{ij})$ as lender i's demand function for data on households on island 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. 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.[41] 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}.[42] This can be written as

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

[39] The DC issuer can market the credit quality data of revealed households, but not 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 $1 - \theta$, opening the model to mimicking strategies, and 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 unrevealed DC users). However, a minor addition to the model would restore our baseline setup: allowing households to open *unused* DC accounts. Opening a DC account and not using it would directly sort a household into the pool of unrevealed DC users. All cash users would do so, as it would come without privacy costs but with the advantage to be pooled with unrevealed DC users that have a higher average credit quality (see Section 4). This implies the same single pool of unrevealed households as in our baseline model and all results would be identical.

[40] The numbering of lenders by i and household island locations by j is symmetric. E.g., Ω_{11} is the data access fee for lender 1 with respect to households that have lender 1 as their 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 home lender.

[41] This becomes clear from the timing of the game laid out in Section 3.5.

[42] We abstract from costs associated with setting up or managing a DC here. The baseline model also abstracts from monetary incentives that the DC issuer could offer the households. The possibility that the DC issuer could, in addition, charge a fee or offer a subsidy on DC use is considered in Appendix B.4.

3.3 Lenders

Lenders in our model face an infinitely elastic supply of funding at cost $c \geq 1$ and engage in Bertrand competition for loans to households. 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.[43] 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, leaving other lenders a pool of lower credit quality borrowers.

We can express lender i's expected profits in general form (with closed forms following as part of the derivations in Section 4):

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

Lender i's profit is given by three terms. The first concerns the expected profits that the lender makes on revealed borrowers whose credit type it has obtained. A household, k, whose credit quality data the lender has obtained, and who has chosen to borrow from lender i, is denoted by ki. A lender's expected profit on a loan with customized loan rate $R_{ki}(q_{ki})$, is $q_{ki} R_{ki}(q_{ki}) - c$.[44] Moreover, m_i represents the mass of revealed households who choose to borrow from lender i. 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 pooled loan rate that lender i charges unrevealed borrowers, and $E[q|u]$ represents the expected credit quality of households conditional on being unrevealed. Here, ui denotes the unrevealed households

[43] In the baseline, the DC is the only source of credit quality data. Appendix B.1 considers alternative sources.

[44] Implicit in this expression is that ki can be represented as a set with continuous support on q_{ki}. The Proof of Proposition 1 shows that this is an accurate representation when solving the game in Section 3.5 by backward induction.

that choose to borrow from lender i. The mass of such borrowers is denoted by υ_i. The third term is the lender's acquisition costs if it chooses to purchase data on revealed borrowers from the DC issuer, as per (2).

3.4 Parameter conditions

We include three conditions, which relate the size of parameters to c, the lenders' cost of funding. First, we let

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

which ensures that the highest quality borrower ($q = 1$) always has a positive NPV project. This is a necessary condition for credit provision to take place. Second, we let

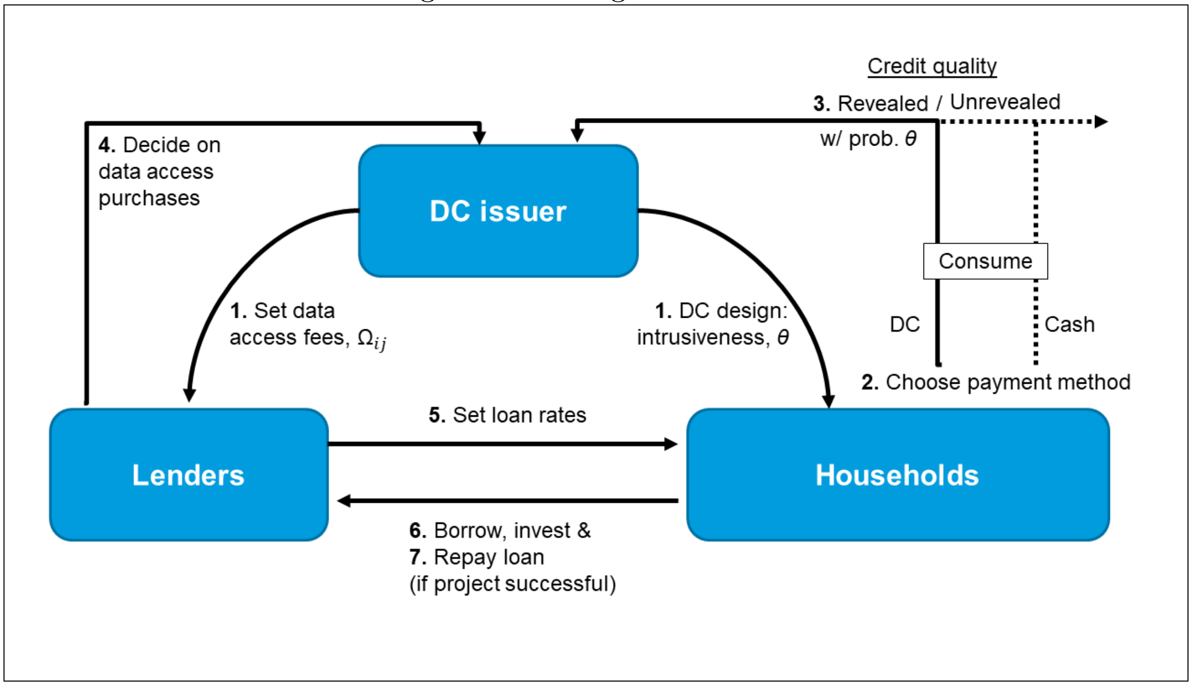
$$\tau < \frac{1}{3}c \tag{5}$$

which means that, when not funded by the home lender (that is, when the household borrows from a bank not located on the same island), a project's value decreases by at most $\frac{1}{3}$ of lenders' funding cost. The reason for this condition is that key mechanisms of our model rely on the competition between lenders and when home lender market power becomes too strong, such mechanisms break down.[45] Third, we assume that

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

which is necessary and sufficient to ensure that households with the highest credit quality ($q = 1$) will always want to disclose their type. This helps preserve tractability.[46]

Figure 1: Timing and actions.



3.5 Timing

The timing of agents' actions and the realization of events is shown in Figure 1. In Stage 1, households are born with an endowment and project opportunity, and the DC provider chooses the DC design, θ, and data access fees, Ω_{ij}. In Stage 2, households observe θ and decide whether to use DC or cash for their consumption and consume, after which DC users' creditworthiness becomes revealed with probability θ at Stage 3. In Stage 4, lenders decide on data access purchases and, subsequently, each lender announces its loan rates in Stage 5.[47] Based on these loan rates, in turn, households decide from which lender to borrow to finance their projects (Stage 6), after which project returns materialize and households with successful projects repay lenders (Stage 7). The next section works backward through the stages of the game to derive the equilibria and elaborates on the optimal strategies of households, lenders and the DC issuer in those equilibria.

[45] See the Proof of Proposition 1.

[46] See the discussion on p.51 for details. Relaxing (6) has been investigated, but leads to highly complex expressions that do not readily lend themselves to analysis.

[47] Namely, individualized loan rates for revealed households on which data has been purchased and a single loan rate for other households.

4 Equilibria

Below we derive the subgame-perfect equilibrium of the game laid out in Section 3.5 in three separate propositions that relate to different credit market outcomes. These propositions are defined for three cases that are delineated by the parameter y and summarized in Table 1. When y is well above c, households will have positive NPV projects ($qy > c$) unless their probability of success q is very low. Referring to borrowers with positive (negative) NPV projects as good (bad) borrowers, the share of bad borrowers is small when y is high. The equilibrium that ensues from Table 1's "High y" case is discussed in Section 4.1. Conversely, when the payoff on a successful project is only barely above the lender's cost of funding, then households will have negative NPV projects ($qy < c$) unless their credit quality is high, leading to a large share of bad borrowers on the loan market. Section 4.2 analyzes the outcomes of Table 1's "Low y" case. The "Intermediate y" case is a direct extension of the other two cases and offers little additional insight: its equilibrium is derived and discussed in Appendix B.7.

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	$y \in \left(c, \frac{4}{3}c\right)$

4.1 Equilibrium with few bad borrowers

When the share of bad borrowers is small, the following equilibrium ensues:

Proposition 1 *When $y > 2c + \tau$, the DC issuer sets $\theta = 1$, charges each lender data access fees $\Omega_{ii} > 0$ for information about its on-island households, with the closed form solution given in (28), and offers each lender free information about off-island households ($\Omega_{ij} = 0$*

when $j \neq i$). Households sort into DC use if

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

and choose cash otherwise, where the closed-form for $E\left[q|\, u\right]$ is displayed in (26). Home lenders buy data access and offer differentiated loan rates

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

to revealed borrowers and a single loan rate

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

to unrevealed borrowers.[48] *All households borrow from their respective home lenders.*

Proof. See Appendix A (p.45). ■

4.1.1 Households' optimal strategies

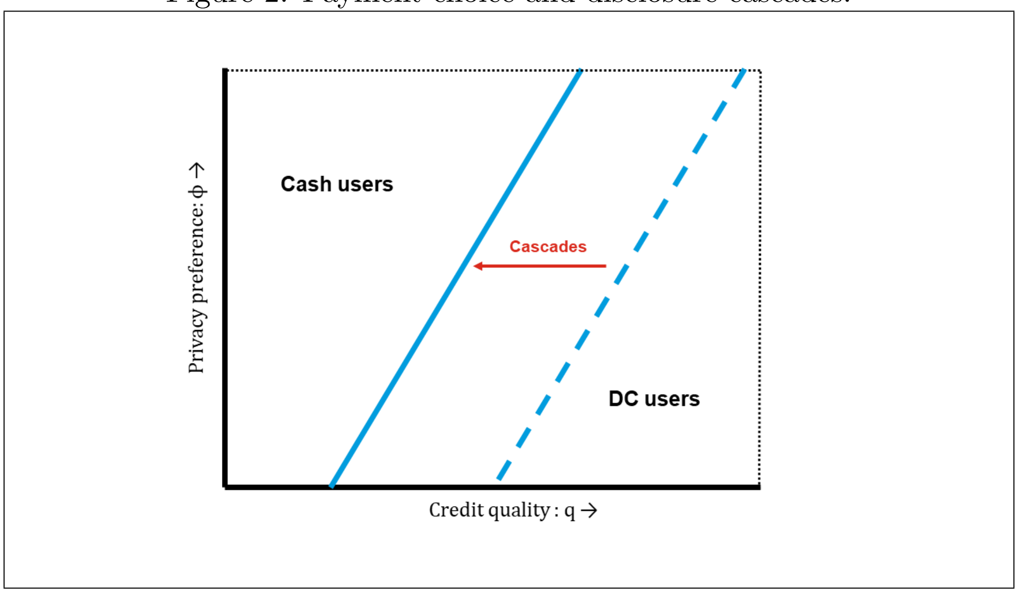
The sorting condition in (7) implies that a household's payment choice is affected by other households' payment choices. A comparison to a cash-only (or, equivalently, $\theta = 0$) setting is useful to highlight these interactions. Without DC, all households would use cash so that $E\left[q|\, u\right] = E\left[q\right] = \frac{3}{4}$ and $R_u = \frac{4}{3}c+\tau$. At this loan rate, some households will find it attractive to use a $\theta > 0$ DC, when given that option. 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 lowers $E\left[q|\, u\right]$ and raises R_u, as some of them become revealed. A lower $E\left[q|\, u\right]$ implies that more households find that the condition in (7) holds, and choose DC over cash. These households are near

[48]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 discussed in the Proof of Proposition 1. We also note that the closed-form for R_u is displayed in (29) and and its comparative statics to the underlying parameters are shown in (30)

(to the left of) the threshold given by the dashed line, meaning that their credit quality is high among cash users. Their choice to use the DC in turn further lowers $E[q|u]$, and so on.

The equilibrium in Proposition 1 documents 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 disclosure: households that only choose to use the DC because other, higher credit quality households choose to do so. Households in between the dashed and unbroken lines Figure 2 can be worse off than if the DC did not exist (i.e., in the cash-only economy), which highlights the public good nature of privacy here. The implications for individual and aggregate welfare are further explored in Section 6.[49]

Figure 2: Payment choice and disclosure cascades.



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 open.

[49] For welfare analysis with information cascades, see also Cong and Xiao (2024).

4.1.2 Lenders' optimal strategies

Lenders engage in Bertrand competition on loan rates, but because of home lender advantage, households are willing to pay a premium up to τ on loans from home lenders and therefore equations (8) and (9) emerge from limit pricing. However, these equilibrium equations are the outcome of not only the pricing game between the lenders, but also the information acquisition game among them.

4.1.3 The DC issuer's optimal strategy

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, in equilibrium, each household does end up borrowing from its home lender and that lenders have the maximum incentive to buy information from the DC issuer. This is attained by maximizing the difference in expected profits between an informed and an uninformed home lender (that is, between a home lender that buys the data and one that does not).

First, we note that lenders that engage in symmetric Bertrand competition always break even. As discussed in the Proof of Proposition 1, this implies that absent some ability for the DC issuer to commit not to sell data about the same borrowers to multiple lenders (further discussed in Section 5), data about off-island borrowers has zero value to each lender. That is, in a symmetric information game, limit pricing prevails and lenders make zero profits on off-island households. Second, the difference between the expected profits that a lender can obtain on its own market (from on-island borrowers) when informed and when not, increases in the degree of adverse selection the lender would suffer when not informed. Therefore, by providing data on off-island borrowers for free, the DC issuer maximizes the potential adverse selection for an uninformed home lender and hence the price it can charge for information about on-island borrowers. Should a lender choose not to buy the data on its on-island revealed households, it would have to offer a single loan rate below (9) to attract some revealed borrowers.[50] In doing so, the uninformed lender would earn less on unrevealed borrowers, while also facing adverse selection on the revealed-household market. A single loan rate that attracts the best borrowers would be below the breakeven rate for the loan portfolio. Overall, cutting rates below (9) reduces the home lender's profits. Therefore, the home lender instead chooses to purchase the data from the DC issuer as long as it makes any positive profit on revealed households and, foreseeing this, the DC issuer charges it a fee marginally below its total earnings on revealed households.[51]

To maximize this fee, the DC issuer sets $\theta = 1$, which makes the set of revealed households as large as possible. A higher θ not only increases the odds of revelation per DC user but also increases the mass of DC users through cascade effects. 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). When θ increases, the mass of unrevealed DC users shrinks, $E\left[q|\, u\right]$ declines and this induces more households to opt for DC use.[52]

4.2 Equilibrium with many bad borrowers

Proposition 2 *When $y \in \left(c, \frac{4}{3}c\right)$, then if $\tau > y - c$, no credit provision takes place and all households use cash. Instead, when $\tau < y - c$, households with $q \geq \frac{\alpha\varphi + c}{y - \tau}$ choose to use the DC and, if revealed, borrow from their home lender with loan rates given by (8). The DC issuer optimally sets $\theta = 1$ and charges lenders a positive data access fee for their on-island borrowers, shown in (36), while offering each lender free data on its off-island borrowers.*

Proof. See Appendix A (p.54). ■

[50] If the home lender offered the loan rate in (9), no revealed household would borrow 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.

[51] This is τ times the mass of revealed households times their expected quality ($\tau m E\left[q|\, dc\right]$, solved in closed form in (28)).

[52] Appendix B.3 shows that when privacy costs are quadratic instead of linear, the mass of DC users may decline as θ increases (depending on paramaters) but $\theta = 1$ nonetheless remains optimal for the DC issuer.

4.2.1 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 low) and there is no viable breakeven loan rate to the unrevealed: households can at most pay y on a successful project, but $E\left[q|\, u\right]y - c < 0$ when $y \in \left(c, \frac{4}{3}c\right)$.[53] Therefore, no lender (including the home lender) could break even on lending to such households. When the loan market for the unrevealed 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 cash, because cash users do not obtain loans. Thus, there are no negative disclosure externalities.

4.2.2 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 attractive enough to induce a choice for DC over cash.

Instead, if $\tau > y - c$, lenders effectively become monopolists: 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 monopoly of the home lender actually leads to a credit market freeze in this setting. At Stage 5 of the game, the home lender will charge revealed households $R_k\left(q\right) = y$, which transfers the full profit of a successful project to the lender. Anticipating this at Stage 2, no household will choose to use the DC.[54]

As the key insights of this paper center on the interaction between DC use and credit provision, we center attention on the case where the home-lender advantage is small enough

[53] This follows from $\sup E\left[q|\, u\right] = E\left[q\right] = \frac{3}{4}$.

[54] This is because there is a privacy cost to using the DC, but no benefit: the DC enables credit access that however yields a zero return.

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 $y < \frac{4}{3}c$.[55]

4.2.3 The DC issuer's optimal strategy

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. 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 between 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.

5 Data usage strategies

So far, we have focused on a DC issuer that markets its data to lenders, using non-exclusive contracts whereby one lender taking up the contract does not preclude other lenders from obtaining access to the same data. Section 5.1 looks at the case where instead the DC issuer is the provider of credit to households. The insights of this case extend to a setup wherein the DC issuer instead offers contracts with exclusive data access to lenders, so that one lender becomes the sole proprietor of part of the data, discussed in Section 5.2. The latter

[55] 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.

setup can equivalently be interpreted as the DC issuer creating its own lending subsidiary and providing only that subsidiary access to its trove of data.[56]

5.1 DC issuer - lender

Figure 3 lays out the timing of a game between households and a monopoly that both issues the DC and provides credit.[57] DC design and household sorting into means of payment comprise the first two stages, after which the monopoly sets loan rates in the third stage and households borrow, invest and (if their projects are successful) repay loans in the final stage.

This setup leads to a stark outcome: all households choose cash and the monopoly makes no profit. At the third stage, the monopoly sets loan rates that seize the expected value added of a revealed borrower's project, $R_k(q) = y$. Foreseeing this at the second stage, households see no benefit to choosing DC, which comes with a privacy cost but no potential benefit from lower loan rates when revealed, and therefore opt for cash.

Proposition 3 *A single monopoly of both payment data and credit provision fails to make profit, as households' outside option of cash leads to zero credit provision when there is too much market power in lending.*

Proof. See Stage 2 in the proof of Proposition 2 in Appendix A.[58] ■

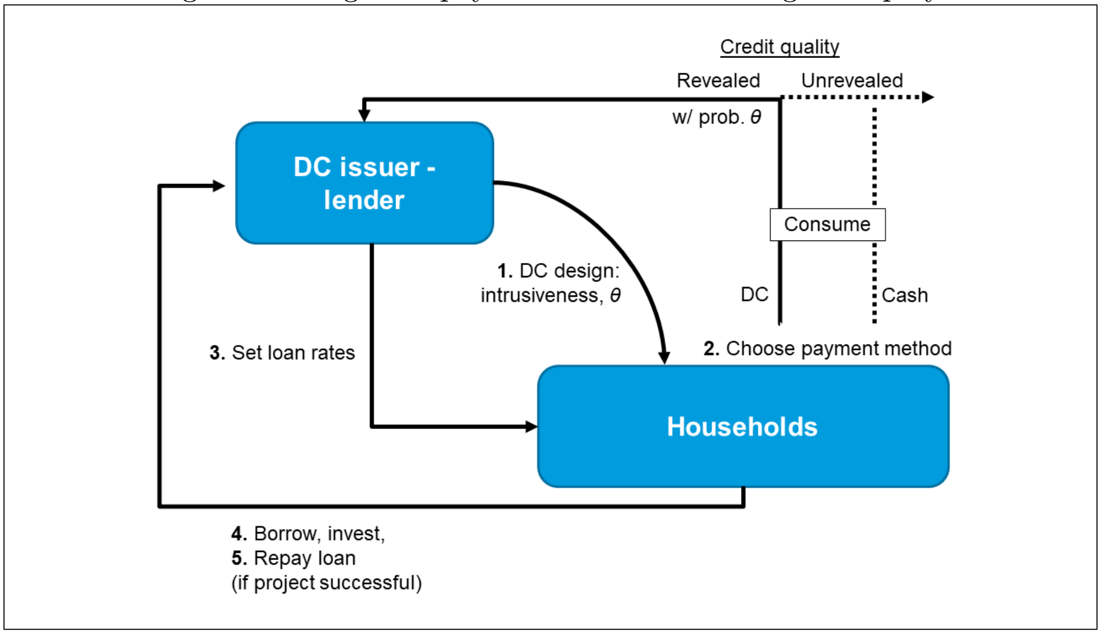
The fundamental problem for the DC issuer here is the ability of households to opt out before being charged monopolistic loan rates. Unmodeled factors, like economies of scope, may plead for the integration of data collection and credit provision. Nevertheless, the notion that there is a benefit for the data monopolist to be a "puppet master", who manipulates information asymmetries from above but does not directly get involved in credit markets, is

[56] This equivalence emanates from profit transfers within the corporate sector having no impact on household sorting and credit provision.

[57] We assume that this monopoly is also the home lender towards all households here: the return on a successful project is y when borrowing from the monopoly.

[58] I.e., formally, Proposition 3 is a direct extension of the part of the proof of Proposition 2, which shows that no credit provision takes place when $\tau > y - c$, as home lenders become de facto lending monopolies for τ large enough and all households use cash.

Figure 3: Integrated payment data and lending monopoly.



powerful. So is the absence of a fully credible way to precommit not to exploit the market power inherent in a data monopoly, after households have given up their data.[59] Market power in lending is a double edged sword in a setting like this: it could facilitate higher loan revenues, but only if enough households stay put and the market power does not compel too many of them to opt out of disclosure altogether.

5.2 Exclusive data access

This section considers exclusive access whereby only one lender (e.g., the highest bidder) is allowed to buy access. Unlike the DC issuer-lender in Section 5.1, lenders with exclusive data access are not monopolists towards revealed households, because such households can still choose to borrow from uninformed lenders too. But that choice is not an attractive one and DC use now unwinds entirely. After obtaining exclusive data access, a lender (regardless of whether it is a home lender or not) optimally charges all revealed households (ε below) the equilibrium loan rate for the unrevealed, R_u, given by (9). This is optimal because the

[59]See Liu et al. (2023a) for empirical evidence on this from Datarade, one of the largest marketplaces for the purchase and sale of data.

revealed households are cornered at this loan rate: they have nowhere to go to obtain the lower loan rates that would be commensurate with their credit qualities, all of which are above the average of the unrevealed pool.[60] Foreseeing that using DC comes with a privacy cost but the same loan rate as using cash, no household sees a benefit to using DC. All households therefore opt for cash, implying zero profit for the DC issuer.[61]

6 Policy analysis

We now draw insights from our model for the analysis and design of socially optimal policy. Section 6.1 considers the role of regulation to constrain the DC's intrusiveness and Section 6.2 analyzes whether access to a second policy instrument, implemented as a tax or subsidy on DC use, helps attain higher welfare.[62]

6.1 Welfare analysis: regulation

To enable the welfare analysis, we extend our model by adding two more stages to the game described in Section 3.5. First, we assume that the policy maker determines regulation at the beginning of the game.[63] We denote socially optimal DC design by θ^* and implement regulation as a constraint on the intrusiveness of data collection.[64] That is, 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: $\theta = \theta^*$.[65] Second, we let y represent an uncertain state

[60] This is shown in Stage 4 of the Proof of Proposition 1.

[61] The above applies to Table 1's high y case. In the low y case, adverse selection precludes uninformed lenders from participating in the credit market. The informed lender charges all revealed households $R_k(q) = y$ at Stage 3 and the outcome is identical to Proposition 3.

[62] In addition, Appendix B.5 discusses the potential use of monetary policy, which can affect lenders' funding cost and Appendix B.6 considers a setting where policy has made household data portable.

[63] The implicit assumption is that regulatory policies on DC would be infrequently revised due to, e.g., practical difficulties in changing data access and commercialization rights ex-post or the significant upfront costs associated with widespread DC adoption, which would necessitate a degree of regulatory certainty.

[64] An example of constraining intrusiveness could be not allowing a BigTech to pair (all of) its social media or ecommerce data with the collected payment data. For example, China has imposed regulatory bounds on the use of payment and platform data in BigTech credit provision.

[65] Alternatively, the policy maker could forbid private DC and instead introduce a CBDC with design θ^*. For further discussion, see the closing paragraph of Section 8.

variable, which realizes in a third stage, after the policy maker's and DC issuer's decision stages: both DC policy and DC design are set long-term, before the aggregate state is known.[66] With probability $\gamma \in [0,1]$, y takes the value y_h and the good economic state of Proposition 1 materializes. With probability $(1-\gamma)$, a bad state, y_l, occurs, in which credit risks are profound and there are relatively many bad borrowers, as in Proposition 2. That is, we let y_h represent "High y" in Table 1, with $y_h > 2c + \tau$, while y_l represents "Low y" in Table 1 with $y_l \in \left(c, \frac{4}{3}c\right)$.[67]

6.1.1 Welfare expression

We define aggregate social welfare as the sum of the expected payoffs of all households and the total profits of lenders and the DC issuer. Welfare is then determined by two factors only: the value added, $qy - c$, of the projects that receive funding;[68] and the privacy costs experienced by DC users. In general form, welfare can be expressed as:

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

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, $E[W]$, with respect to θ:

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

$$\begin{aligned}
\max_\theta E[W] &= \max_\theta \{\gamma W_h + (1-\gamma) W_l\} && (12)\\
W_h &= \frac{3}{4} y_h - c - 2\alpha\theta \left[1 - \left(1 + \frac{4\alpha}{3c}\right) E[q|\, u]\right] && (13)\\
W_l &= \theta\lambda && (14)
\end{aligned}$$

[66] The remaining stages (Stages 2-7 in Section 3.5) become Stages 4-9.

[67] Focusing on the "High y" and "Low y" cases in Table 1 permits analytical solutions. Over the intermediate range $y \in \left(\frac{4}{3}c, 2c+\tau\right)$, welfare can be found numerically, as discussed in Appendix B.7.

[68] From Propositions 1 and 2, in equilibrium households borrow from home lenders and therefore return conditional on project success is y, so that a project's expected NPV is $qy - c$. 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.

where W_h and W_l represent welfare in the good and bad states, respectively, the closed-form for $E[q|u]$ is in (26), and λ is a collection of constants, with $\lambda > 0$, shown in (40).

Proof. See Appendix A (p.56). ■

6.1.2 Socially optimal regulation

Figure 4 (on p.30) shows how θ impacts W_h and W_l. In the bad state, welfare increases linearly as θ increases. A higher θ helps households who want to reveal their types to do so, while means of payment choice is purely individual and there are no disclosure externalities when the loan market for unrevealed households is inoperative. Instead, in the good state, all households receive loans and therefore the total value added of projects is independent of θ. The DC's impact on aggregate welfare then runs through privacy costs. A higher θ has a direct effect on privacy costs in (1) and in addition has an indirect effect from disclosure cascades. These effects are seen from, respectively, the terms θ and $E[q|u]$ in (13). Put together, direct and indirect effects imply that W_h decreases more than linearly as θ increases.[69] This lays the foundation for an interior solution for socially optimal policy, derived in Proposition 5.

Proposition 5 *When the good state is certain ($\gamma = 1$), privacy costs dominate and the policy maker bans the DC, $\theta^* = 0$. When the bad state is certain ($\gamma = 0$), credit inclusion from type differentiation dominates and laissez-faire, $\theta^* = 1$, is socially optimal. When $\gamma \in (0,1)$, the policy maker for some parameterizations chooses to regulate the DC with $\theta^* \in (0,1)$, solved in (41).*

Proof. See Appendix A (p.58). ■

[69] There are two reasons for this nonlinearity. 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 (26) and footnote 103, $E[q|u]$ decreases convexly as θ rises. 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.

6.1.3 Distributional effects

The distributional effects from DC design differ between the two states of y. In the bad state, DC users' disclosure is a Pareto improvement, creating value to them, no externalities to others, and profits to the corporate sector. In the good state, some (high q, low φ) households gain from lower customized loan rates. But their gains are exactly offset by the losses of unrevealed households, who face a higher loan rate.[70] Moreover, total corporate profits do not depend on θ in the good state: a higher θ transfers profits within the corporate sector, as the DC issuer gains data access fee revenues at lenders' expense. The impact of θ on W_h therefore derives from the one effect that has no offset: the privacy loss of some DC users that were swept along in disclosure cascades.

6.1.4 Comparative statics

The solution for θ^* in (41) enables an analysis of the comparative statics of optimal regulation to underlying parameters. We highlight three comparative statics of particular interest.

Society's care for privacy In (1), α represents the weight that society places on privacy preferences. The main insight from varying α is summarized in Figure 5: laissez-faire is optimal in the extremes where society cares not at all or a great deal about privacy. It is when society cares somewhat but not too much about privacy that there is a role for regulation to optimally constrain DC intrusiveness:

Corollary 1 (to Proposition 5) *Regulation has a role to play ($\theta^* < 1$) when the two means of payment each have enough users.*

For $\alpha \to 0$, DC use and thereby disclosure is maximized, but privacy ceases to matter. Expected welfare in (12) becomes $E[W] = \gamma\left(\frac{3}{4}y_h - c\right) + (1-\gamma)\,\theta\lambda$, which monotonically

[70] In practice, such distributional effects could materialize along socially undesirable dimensions, such as race or gender (Chen et al., 2023; Fuster et al., 2022). For estimates of the effect of the increased availability of borrower data on welfare and the distribution of gains and losses, see Blattner et al. (2022) and Jansen et al. (2022).

increases as θ rises, implying $\theta^* = 1$. Instead, when α is very high, cash use dominates. In this case too, $\theta^* = 1$ can be observed for some parameterizations.[71] When society cares a great deal about privacy, it is, in effect, "doing the regulator's job" in terms of internalizing privacy costs. Disclosure cascades are limited, because the base of DC users from which such effects emanate is small. However, when society has an intermediate care for privacy, cascades can both affect many households and be painful, and θ^* can become interior or even 0, as per Proposition 5.

Market power in lending The relation between market power and optimal regulation centers on the bad state, where $\frac{\partial \theta^*}{\partial \tau} < 0$. Welfare in the good state is not affected by market power: τ does not enter W_h in (13).[72] Instead, in the bad state, a higher τ increases the loan rate of DC users, but does not affect cash users, who do not obtain credit. A higher τ then pushes more households into cash use. With fewer DC users in the bad state, the social benefits of a larger θ in the bad state decline, because these stem from DC users who can become revealed and thereby obtain credit.[73] Thus, θ^* decreases.

The type of economy A potentially interesting question to ask is what parameters in our model might characterize a developing relative to an advanced economy. In many developing economies, a sizable part of the population is financially excluded and cannot access credit. The closest our model can come to matching this, is in expected terms: when γ is smaller, the expected mass of households without credit access is larger. From (12), $\frac{\partial \theta^*}{\partial (1-\gamma)} > 0$: when the bad state becomes more likely, the policy maker focuses more on the gains from type differentiation for credit provision. From this angle, optimal DC regulation for a devel-

[71] As α enters the term for θ^* in (41) many times, an analytical derivation of this comparative static is challenging to obtain. Instead, numerical investigations (available on request) highlight that $\theta^* = 1$ can for some parameterizations be observed for α large enough (but still satisfying (6)).

[72] The reason is that all households get credit in the good state and both unrevealed and revealed households pay τ as part of the equilibrium loan rates, (8) and (9). That is, in the good state τ does not drive a wedge between cash and DC using households and therefore does not affect the policy maker's considerations. A higher τ leads to lower payoffs to households in (1), but from an aggregate welfare perspective these are exactly offset by higher lender profits and therefore $\frac{\partial W_h}{\partial \tau} = 0$.

[73] Formally, $\frac{\partial W_l}{\partial \theta} = \lambda$ and $\frac{\partial \lambda}{\partial \tau} < 0$ from (40).

Figure 4: Welfare in the good and bad states.

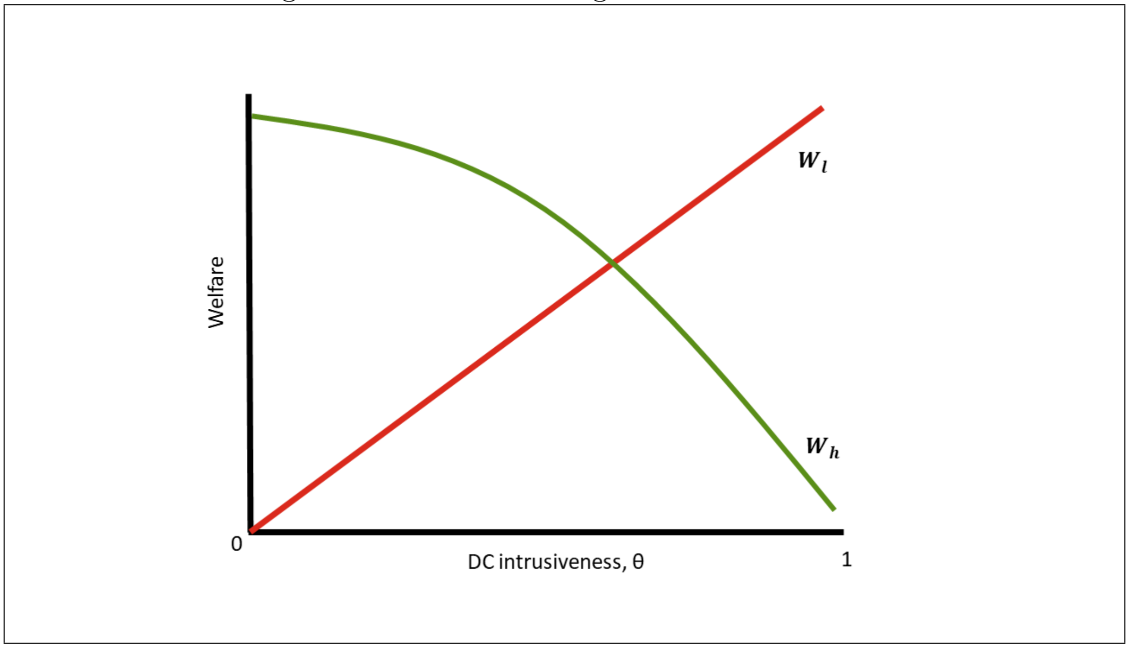
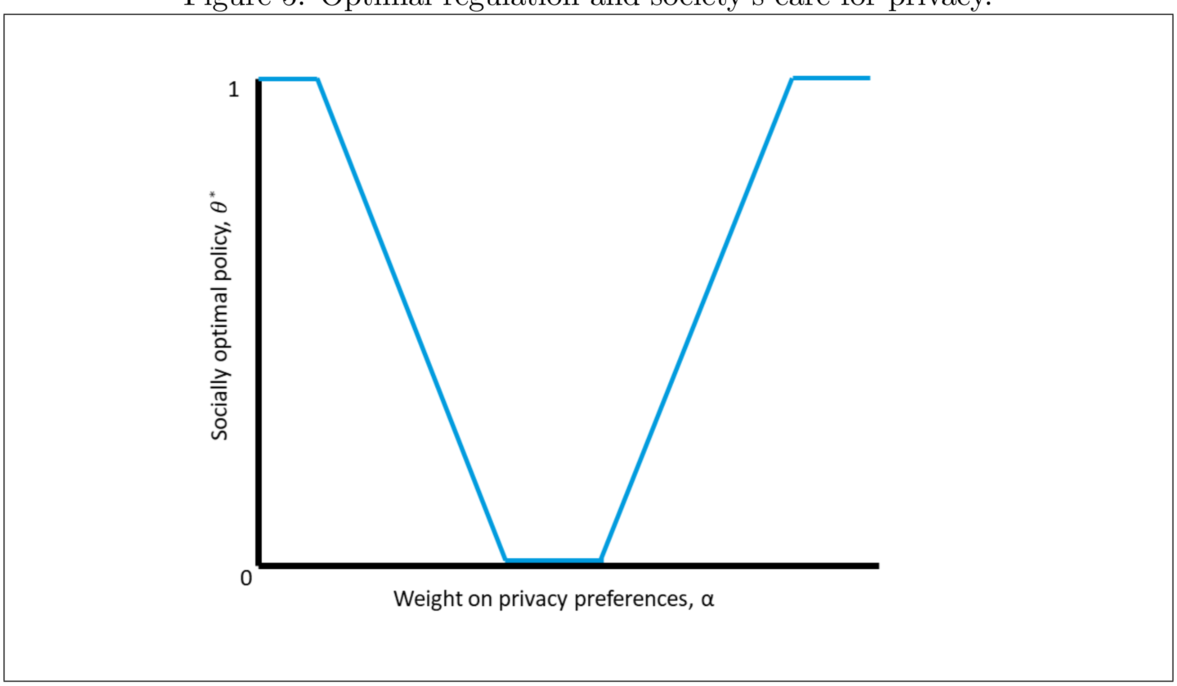


Figure 5: Optimal regulation and society's care for privacy.



oping economy might be looser than for an advanced economy. However, another facet of developing economies is that they tend to face greater economic volatility. A stylized way to represent that in our framework is a "mean preserving spread": equally increasing y_h and decreasing y_l. Widening such a spread implies a lower θ^* and tighter DC regulation.[74]

6.2 A second instrument

Regulation affects how much household data is disclosed, which is what a social planner cares about in our framework. Household disclosure consists of two margins: who sorts into DC use, which we will refer to as the extensive margin of disclosure; and how much information is disclosed about DC-using households, which we term the intensive margin. By constraining the intensive margin, regulation also affects sorting into means of payment and thereby the extensive margin. But the fact that there are two margins and one instrument implies that there could be room for a second instrument to improve welfare outcomes.

We here explore such a second instrument in the form of a government tax or subsidy on DC use.[75] We denote this instrument by T, where $T > 0$ represents a tax and $T < 0$ is a subsidy.[76] Unlike the regulatory constraint, the tax (subsidy) cannot target the intensive margin, but it can affect the extensive margin by disincentivizing (incentivizing) household sorting into DC use. We rederive the model with the inclusion of T and the resulting expressions for the welfare analysis are recorded in Appendix B.8.

Proposition 6 *By combining regulation with either a tax or a subsidy on DC use, optimal policy can attain higher expected welfare than regulation alone.*

Proof. In text (below). ■

[74] Although y_h enters W_h in (13), it does not interact with θ in that term: a higher return in good times is an equivalent boon for all, since everyone receives credit. Instead, in bad times, a lower return means fewer gains from credit quality differentiation and therefore the policy maker's tradeoff shifts to a lower θ^*.

[75] This could be in the form of a tax/subsidy on, e.g., DC transactions or opening a user wallet.

[76] An implicit assumption is that reduced (increased) DC-based consumption from taxation (subsidization) is assumed not to affect the DC issuer's abilty to infer credit quality data on a given household.

The optimal use of two instruments always weakly improves welfare, because $T = 0$ can replicate the regulation-only outcome. For some parameterizations, the availability of a second instrument does not raise welfare.[77] For other parameterizations, it does: Figures 6 and 7 provide numerical examples.[78] Figures 6 and 7 draw expected welfare against θ for given values of T. In particular, they show the $T = 0$ case (unbroken line) that represents regulation acting alone, as well as the case of socially optimal policy with two tools: the peak of the dashed line attains the highest expected welfare when using both tools.[79]

In Figure 6, the good state is very likely ($\gamma = 0.95$) and therefore the costs of intrusiveness in that state are a relatively dominant consideration. Regulation on its own would set $\theta^* = 0$, banning DC use. The downside of doing so, is that when the bad state occurs, no credit provision takes place. A tax now allows the policy maker to exit this corner.[80]

Figure 7 uses the same parameterization as Figure 6, except that both states are equally likely ($\gamma = 0.5$). The policy maker now optimally introduces a subsidy. In the bad state, households disclose less data than socially optimal, even when $\theta = 1$. This derives from lender market power, τ, which prevents some positive NPV households from disclosing in the bad state (Section 6.1.4). A subsidy leans against this underdisclosure. With more DC users in the bad state, the tradeoff for optimal regulation now shifts to give more weight to that state and the subsidy is paired with looser regulation. Figures 6 and 7 highlight that the policy maker sometimes optimally applies the two instruments as substitutes (Figure 6) and sometimes as complements (Figure 7).

[77] E.g., in a corner where zero DC uptake is socially optimal, $\theta^* = 0$ is as effective as an infinite tax or a combination of both.

[78] The parameterization behind these figures is $y_h = 2.2$, $y_h = 1.3$, $c = 1$, $\alpha = 0.05$ and $\tau = 0.1$ (which satisfies the relevant parameter constraints) while γ is varied. We also note that the expressions in Appendix B.8 do not readily lend themselves to an analytical derivation of combined optimal policy.

[79] This social optimum is found through a grid search.

[80] Setting $T = 0.1$ and $\theta^* = 0.14$, the policy maker permits a limited degree of DC use and therefore differentiation and credit provision in the bad state (among the highest quality borrowers that are willing to pay the tax, given their high disclosure benefits).

Figure 6: Optimal combination of regulation and a tax.

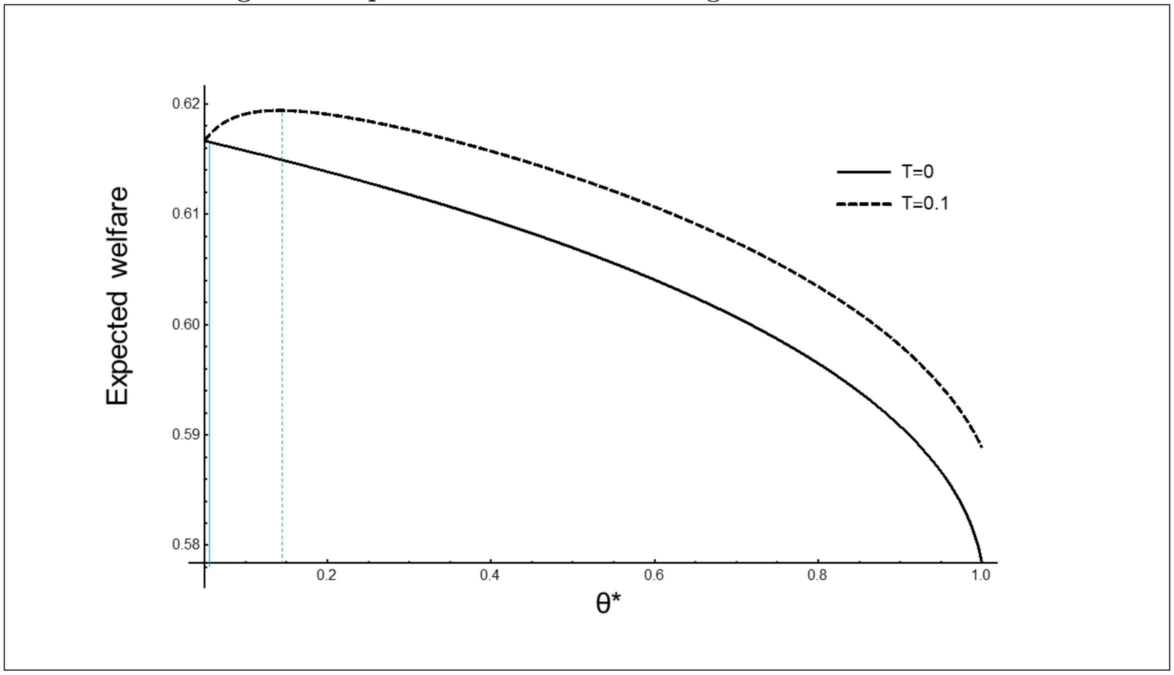
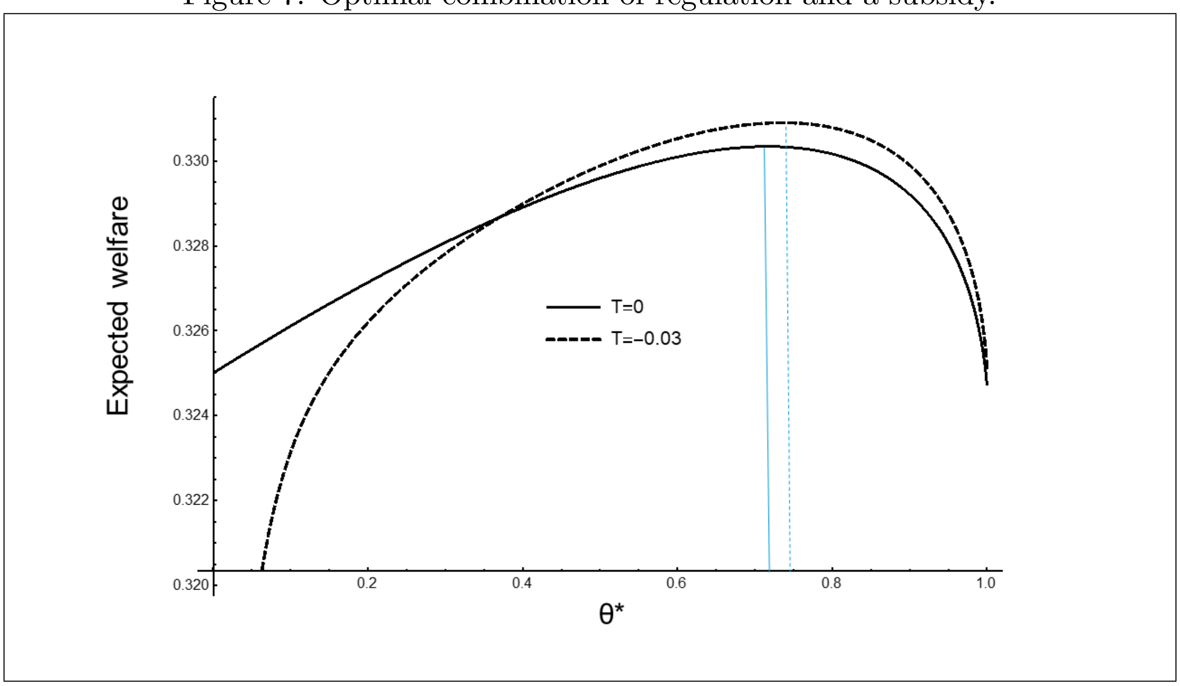


Figure 7: Optimal combination of regulation and a subsidy.



7 Lender consortium

The baseline model centers on a DC issuer that exploits information asymmetries on the credit market. We here develop a comparison to a case wherein the lenders own the DC: the DC is set up by a consortium of lenders, which decides on θ. We assume that the collected DC data is now freely available to all lenders.[81] While lenders cooperate in DC design, their mode of competition on the credit market is unchanged compared to the baseline.[82]

Without data access fees, a lender's total profit is always $\frac{3}{4}\tau$ in the good state and therefore its profits do not depend on θ.[83] Hence, the lender consortium is indifferent about θ in this case and privately optimal θ is indeterminate. Instead, in the bad state, only revealed households receive credit and the mass of such households rises as θ increases, as can be seen from (32) and (34). In equilibrium, lenders expect to earn $q\left(\frac{c}{q}+\tau\right)-c=q\tau$ per revealed on-island borrower and, since there are more such borrowers when θ is higher, lender profits depend positively on θ and the consortium sets $\theta=1$.[84]

In sum, the difference between the independent DC issuer and the DC owned by a lender consortium centers on the good state only. In that state, the DC issuer maximizes intrusiveness to induce more households to use the DC, which gives it a better hand to play towards the lenders. This is a private gain for the DC issuer and is therefore not replicated by the consortium. Conversely, $\theta=1$ in the bad state creates value for households and lenders, and is optimal for DC issuer, lender consortium and social planner alike.

Proposition 7 *Unlike the DC issuer, the lender consortium does not actively exploit dis-*

[81] Data access fees would not make a difference here if consortium profits are remitted back to the lenders. because such fees act as a net transfer within the corporate sector.

[82] For empirical evidence on lender cooperation through a private credit bureau, see Liberti et al. (2022).

[83] In equilibrium, every lender expects to earns $E[q|u]\left(\frac{c}{E[q|u]}+\tau\right)-c=E[q|u]\tau$ on an unrevealed on-island borrower and $q\left(\frac{c}{q}+\tau\right)-c=q\tau$ on a revealed on-island borrower. This lender's total earnings are then $(1-m)E[q|u]\tau+mE[q|dc]\tau$ where m is the mass of revealed households. Since $E[q]=(1-m)E[q|u]+mE[q|dc]$, total earnings simplify to $E[q]\tau$. Moreover, $E[q]=\frac{3}{4}$ and therefore a lender earns $\frac{3}{4}\tau$ in the good state.

[84] Considering the setup with an uncertain state realization of Section 6, the lender consortium chooses $\theta=1$, because with probability γ the good state occurs and the consortium is indifferent about θ and with probability $1-\gamma$ the bad state comes about and the consortium prefers $\theta=1$.

closure externalities in the good state; unlike a social planner, the consortium does not lean against such externalities either: optimal θ is indeterminate in the good state.

Proof. In text (above). ■

8 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, appear poised to expand upon their already vast trove of knowledge about their customers and to develop 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 (Haksar et al., 2021).

This paper builds a framework to analyze how private and social optima diverge when a monopolist offers privacy-valuing households a digital payment system that collects data on their habits but can help them signal their creditworthiness to lenders. We find that the monopoly uses its pole position to play off both households and lenders against each other. It squeezes households into a partially unraveling Lemons' Market, whereby the willingness of some high quality households to disclose prods more tranches of households to differentiate themselves to the detriment of their privacy. To maximize the value of household data in its hands, the monopolist makes its data collection as intrusive as possible, which optimally amplifies the Lemons' Market dynamics. The monopolist optimizes by pricing information to lenders in a way that augments adverse selection for those that choose not to buy data access and thus maximizes the difference in profits between an informed and an uninformed lender. Because households have an outside payment option, the monopolist earns the most as a puppet master who harnesses competition on the loan market by manipulating information asymmetries, rather than getting into the lending fray itself. Giving households ownership of their own data or giving lenders' ownership of the data-collecting technology does not eliminate the externalities that households impose on each other in equilibrium. However, a data monopolist actively amplifies these externalities.

Socially optimal data regulation aims at shielding households from being played off against each other. The challenge in doing so, is that regulation is a long-term policy, while the benefits from counteracting intrusive data collection are state dependent. In a bad state, where projects become less profitable, credit provision to unrevealed borrowers can freeze. The monopoly's data gathering can then be a net positive, not just to a subset of households, but to aggregate welfare, because the data unfreeze part of the credit market by differentiating better quality households. Socially optimal regulation trades off the partial preservation of credit provision during downturns against the privacy losses of households that are induced into disclosure cascades in upturns. Optimal regulation is most interventionist when market power in lending is strong and when the monopolist's payment system and cash both have solid take-up.

Taxing or subsidizing households who use the monopolist's payment system can be a second policy tool. Taxation and subsidization work differently from regulation, because these only move the extensive margin of data collection (who opts in), while regulation also affects the intensive margin (the extent of data collection per participating household). The two policies together can attain higher welfare than regulation alone and depending on, e.g., the likelihood of the good state with data overprovision and the bad state with data underprovision, can be either optimally used as substitutes or as complements.

This paper focuses on a single payment provider and one outside payment option. Future work can build on the framework provided here to consider competition between multiple payment providers, which could include a public provider, such as a CBDC.[85] Indeed, policy discussions on the role of data sharing through CBDC are gaining increased prominence (ECB, 2023). While our model would need extending to speak to competition between CBDC

[85] In particular, the extension in Appendix B.3 may permit the co-existence of multiple, differentiated means of payment that each draw a market share.

and multiple means of payment, it can already be interpreted from the angle of a benchmark case wherein CBDC is introduced with data-sharing features as the sole alternative to cash. What has been discussed as optimal regulation (θ^*) of a private digital payment system, can equivalently be considered the optimal degree of data collection with a welfare-maximizing CBDC.

References

Acemoglu, D., Makhdoumi, A., Malekian, A., and Ozdaglar, A. (2022). Too Much Data: Prices and Inefficiencies in Data Markets. *American Economic Journal: Microeconomics*, 14(4):218–56.

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

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

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.

Allen, F., Carletti, E., and Marquez, R. (2011). Credit Market Competition and Capital Regulation. *The Review of Financial Studies*, 24(4):983–1018.

Allen, F. and Gale, D. (1999). Diversity of Opinion and Financing of New Technologies. *Journal of Financial Intermediation*, 8(1):68–89.

Allen, F., Gu, X., and Jagtiani, J. (2021). A Survey of Fintech Research and Policy Discussion. *Review of Corporate Finance*, 1(3-4):259–339.

Athey, S., Catalini, C., and Tucker, C. (2017). The Digital Privacy Paradox: Small Money, Small Costs, Small Talk. NBER WP 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. SSRN Electronic Journal.

Bank, T., Segev, N., and Shaton, M. (2023). Relationship Banking and Credit Scores: Evidence from a Natural Experiment. Discussion Paper 2023.05, Bank of Israel.

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

Berg, T., Burg, V., Gombović, A., and Puri, M. (2020). On the Rise of FinTechs: Credit Scoring Using Digital Footprints. *Review of Financial Studies*, 33(7):2845–2897.

Berg, T., Fuster, A., and Puri, M. (2022). FinTech Lending. *Annual Review of Financial Economics*, 14(1):187–207.

Bergemann, D. and Bonatti, A. (2023). Data, Competition, and Digital Platforms. Discussion Papers 2343R, Cowles Foundation.

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

Bian, B., Ma, X., and Tang, H. (2023a). The Supply and Demand for Data Privacy: Evidence from Mobile Apps. SSRN Electronic Journal.

Bian, W., Cong, L. W., and Ji, Y. (2023b). The Rise of E-Wallets and Buy-Now-Pay-Later: Payment Competition, Credit Expansion, and Consumer Behavior. WP 31202, NBER.

Bijlsma, M., van der Cruijsen, C., and Jonker, N. (2022). Consumer Willingness to Share Payments Data: Trust for Sale? *Journal of Financial Services Research*, 64:41–80.

Bijlsma, M., van der Cruijsen, C., and Jonker, N. (2023). Not all Data are Created Equal - Data Sharing and Privacy. *Applied Economics, In Press*.

Blattner, L., Hartwig, J., and Nelson, S. (2022). Information Design in Consumer Credit Markets. mimeo.

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

Bond, P. and Zeng, Y. (2022). Silence is Safest: Information Disclosure when the Audience's Preferences are Uncertain. *Journal of Financial Economics*, 145(1):178–193.

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.

Bouvard, M., Casamatta, C., and Xiong, R. (2022). Lending and Monitoring: Big Tech vs Banks. TSE Working Papers 1386, Toulouse School of Economics.

Brits, H. and Jonker, N. (2023). The Use of Financial Apps: Privacy Paradox or Privacy Calculus? WP 794, De Nederlandsche Bank.

Brunnermeier, M. K. and Payne, J. (2022). Platforms, Tokens, and Interoperability. WP 2022-8, Princeton University.

Brunnermeier, M. K. and Payne, J. (2023). Strategic Money and Credit Ledgers. WP 31561, National Bureau of Economic Research.

Budak, J., Rajh, E., and Škrinjarić, B. (2023). Resilience to Online Privacy Violation: Developing a Typology of Consumers. *Sc. Annals of Econ. and Business*, 70(3):379–98.

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

Chen, L., Huang, Y., Ouyang, S., and Xiong, W. (2021). The Data Privacy Paradox and Digital Demand. NBER Working Papers 28854, National Bureau of Economic Research.

Chen, S., Doerr, S., Frost, J., Gambacorta, L., and Shin, H. (2023). The Fintech Gender Gap. *Journal of Financial Intermediation*, 54:101026.

Chiu, J. and Wong, T.-N. (2022). Payments on Digital Platforms: Resiliency, Interoperability and Welfare. *Journal of Economic Dynamics and Control*, 142:104173.

Choe, C., Matsushima, N., and Shekhar, S. (2023). The Bright Side of the GDPR: Welfare-Improving Privacy Management. WP 10617, CESifo.

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.

Choi, S., Kim, B., Kim, Y.-S., and Kwon, O. (2023). Central Bank Digital Currency and Privacy: A Randomized Survey Experiment. WP 1147, Bank for International Settlements.

Cicala, S., Hémous, D., and Olsen, M. (2022). Adverse Selection as a Policy Instrument: Unraveling Climate Change. WP 30283, NBER.

Collis, A., Moehring, A., Sen, A., and Acquisti, A. (2022). Information Frictions and Heterogeneity in Valuations of Personal Data. SSRN Electronic Journal.

Cong, L. W. and Mayer, S. (2023). Antitrust, Regulation, and User Union in the Era of Digital Platforms and Big Data. WP 30881, National Bureau of Economic Research.

Cong, L. W. and Xiao, Y. (2024). Information Cascades and Threshold Implementation: Theory and an Application to Crowdfunding. *The Journal of Finance forthcoming*.

Cunliffe, J. (2023). Money and Payments: A Black Ships Moment? Speech delivered at the Economics of Payments XII Conference at the Federal Reserve Board on Oct. 26, 2023.

Dalton, P. S., Pamuk, H., Ramrattan, R., Uras, B., and van Soest, D. (2023). E-Payment Technology and Business Finance: A Randomized Controlled Trial with Mobile Money. *Management Science forthcoming*.

D'Andrea, A. and Limodio, N. (2023). High-Speed Internet, Financial Technology, and Banking. *Management Science forthcoming*.

DeFusco, A. A., Tang, H., and Yannelis, C. (2022). Measuring the Welfare Cost of Asymmetric Information in Consumer Credit Markets. *Journal of Financial Economics*, 146(3):821–840.

Degryse, H. and Ongena, S. (2005). Distance, Lending Relationships, and Competition. *The Journal of Finance*, 60(1):231–266.

Dekel, I., Cummings, R., Heffetz, O., and Ligett, K. (2023). The Privacy Elasticity of Behavior: Conceptualization and Application. In *Proceedings of the 24th ACM Conference on Economics and Computation*, EC '23, page 516, New York, NY, USA. ACM.

Doerr, S., Frost, J., Gambacorta, L., and Shreeti, V. (2023a). Big Techs in Finance. WP 1129, Bank for International Settlements.

Doerr, S., Gambacorta, L., Guiso, L., and del Villar, M. S. (2023b). Privacy Regulation and Fintech Lending. BIS Working Papers 1088, Bank for International Settlements.

ECB (2021). Eurosystem Report on the Public Consultation on a Digital Euro. Technical report, European Central Bank.

ECB (2023). Digital Euro: The Next Step in the Advancement of Our Currency. Report, European Central Bank.

Engels, C., Francis, B. B., and Philip, D. (2022). The Cost of Privacy Failures: Evidence from Bank Depositors' Reactions to Breaches. SSRN Electronic Journal.

Fainmesser, I. P., Galeotti, A., and Momot, R. (2023). Digital Privacy. *Management Science*, 69(6):3157–3173.

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.

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., and Walther, A. (2022). Predictably Unequal? The Effects of Machine Learning on Credit Markets. *The Journal of Finance*, 77(1):5–47.

Gambacorta, L., Khalil, F., and Parigi, B. M. (2022). Big Techs vs Banks. BIS Working Papers 1037, Bank for International Settlements.

Gambato, J. and Peitz, M. (2023). Platform-Enabled Information Disclosure. Discussion Paper Series 224, CRC TR.

Garratt, R. and Lee, M. J. (2021). Monetizing Privacy. Staff Reports 958, NY Fed.

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. (2022). FinTech Lending and Cashless Payments. SSRN Electronic Journal.

Goldfarb, A. and Que, V. F. (2023). The Economics of Digital Privacy. *Annual Review of Economics*, 15(1):267–286.

Goldfarb, A. and Tucker, C. (2012). Shifts in Privacy Concerns. *American Economic Review*, 102(3):349–53.

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

Guennewig, M. (2022). Currency Competition with Firms. SSRN Electronic Journal.

Haksar, V., Carrière-Swallow, Y., Islam, E., Giddings, A., Kao, K., Kopp, E., and Quirós-Romero, G. (2021). Towards a Global Approach to Data in the Digital Age. Staff Discussion Note 21/14, International Monetary Fund.

Hau, H., Huang, Y., Shan, H., and Sheng, Z. (2019). How FinTech Enters China's Credit Market. *AEA Papers and Proceedings*, 109:60–64.

He, Z., Huang, J., and Zhou, J. (2023). Open Banking: Credit Market Competition When Borrowers Own the Data. *Journal of Financial Economics*, 147(2):449–474.

Hu, B., Liu, Y.-L., and Yan, W. (2023). Should I Scan my Face? The Influence of Perceived Value and Trust on Chinese Users Intention to Use Facial Recognition Payment. *Telematics and Informatics*, 78:101951.

Huang, J. (2021). Fintech Expansion. SSRN Electronic Journal.

Huang, J. (2022). The Economics of Consumer Credit on an E-Commerce Platform. SSRN.

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

Ichihashi, S. (2023). Dynamic Privacy Choices. *American Economic Journal: Microeconomics*, 15(2):1–40.

Jansen, M., Nagel, F., Yannelis, C., and Zhang, A. L. (2022). Data and Welfare in Credit Markets. WP 30235, NBER.

Jiang, E. X., Yu, G. Y., and Zhang, J. (2023). Bank Competition amid Digital Disruption: Implications for Financial Inclusion. SSRN Electronic Journal.

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. WP 29096, National Bureau of Economic Research.

Jones, C. I. and Tonetti, C. (2020). Nonrivalry and the Economics of Data. *American Economic Review*, 110(9):2819–58.

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.

Kahn, C. M. and Roberds, W. (2008). Credit and Identity Theft. *Journal of Monetary Economics*, 55(2):251–264.

Kahn, C. M. and van Oordt, M. R. (2023). The Demand for Programmable Payments. SSRN.

Kim, S. J. (2021). Privacy, Information Acquisition, and Market Competition. SSRN.

Li, J. (2023). Predicting the Demand for Central Bank Digital Currency: A Structural Analysis with Survey Data. *Journal of Monetary Economics*, 134:73–85.

Liberti, J., Sturgess, J., and Sutherland, A. (2022). How Voluntary Information Sharing Systems Form: Evidence from a U.S. Commercial Credit Bureau. *Journal of Financial Economics*, 145(3):827–849.

Lin, T. (2022). Valuing Intrinsic and Instrumental Preferences for Privacy. *Marketing Science*, 41(4):60–64.

Lin, T. and Strulov-Shlain, A. (2023). Choice Architecture, Privacy Valuations, and Selection Bias in Consumer Data. In *Proceedings of the 24th ACM Conference*. ACM.

Liu, E., Ma, S., and Veldkamp, L. (2023a). Data Sales and Data Dilution. mimeo.

Liu, Z., Sockin, M., and Xiong, W. (2023b). Data Privacy and Algorithmic Inequality. Working Paper 31250, National Bureau of Economic Research.

Markovich, S. and Yehezkel, Y. (2023). For the Public Benefit: Data Policy in Platform Markets. SSRN Electronic Journal.

Meyer, J. and Teppa, F. (2023). Euro Area Consumers' Payment Behaviour and Banking Digitalisation. Working Paper 772, De Nederlandsche Bank.

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

Nam, R. J. (2022). Open Banking and Customer Data Sharing: Implications for Fintech Borrowers. SAFE Working Papers 364, Leibniz Institute for Financial Research.

OECD (2023). Data Portability in Open Banking: Privacy and Other Cross-Cutting Issues. Digital Economy Papers 348, Organisation for Economic Cooperation and Development.

Ouyang, S. (2022). Cashless Payment and Financial Inclusion. SSRN Electronic Journal.

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*, 35(11):4985–5024.

Petropoulos, G., Martens, B., Parker, G., and Van Alstyne, M. (2023). Platform Competition and Information Sharing. WP 10663, CESifo.

Prince, J. T. and Wallsten, S. (2022). How Much is Privacy Worth Around the World and Across Platforms? *Journal of Economics and Management Strategy*, 31(4):841–861.

Qi, S., De Haas, R., Ongena, S., Straetmans, S., and Vadasz, T. (2023). Move a Little Closer? Information Sharing and the Spatial Clustering of Bank Branches. WP 223, EBRD.

Rishabh, K. and Schäublin, J. (2021). Payment Fintechs and Debt Enforcement. WWZ Working Paper 2021/02, University of Basel.

Sarkisyan, S. (2023). Instant Payment Systems and Competition for Deposits. SSRN.

Schneider, F. and Taudien, R. (2023). Anonymity of Credit-Based Payment Systems. Mimeo.

Soleymanian, M., Weinberg, C. B., and Zhu, T. (2021). Privacy Concerns, Economic Benefits, and Consumer Decisions: A Multi-Period Panel Study of Consumer Choices in the Automobile Insurance Industry. SSRN Electronic Journal.

Tan, B. J. (2023). Central Bank Digital Currency and Financial Inclusion. IMF Working Paper 23/48, International Monetary Fund.

Tang, H. (2023). The Value of Privacy: Evidence from Online Borrowers. mimeo.

Verdier, M. (2020). Digital Currencies and Bank Competition. SSRN Electronic Journal.

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 applies backward induction to solve for the Subgame Perfect Nash Equilibrium (with symmetric pure strategies among the other lenders). We first note that equations (8) and (9) are considered the limit cases of, respectively, $R_k(q) = \frac{c}{q} + \tau - \varepsilon$ and $R_u = \frac{c}{E[q|u]} + \tau - \varepsilon$ with $\varepsilon \to 0$.[86] Moreover, we save notation by focusing on a single subset of households (as results extend symmetrically to all subsets) with a given home lender, h, and other (non-home) lenders denoted by i. Thus, $R_{kh}(q)$ and $R_{ki}(q)$ are the differentiated loan rates that, respectively, the home lender and other lenders charge revealed households; and R_{uh} and R_{ui} are such loan rates charged to unrevealed households.

Stage 6. This is the last decision stage of the game outlined in Section 3.5. 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 DC in Stage 2 and, with probability θ, became revealed at Stage 3). Given limited liability, as incorporated in (1), households are always willing to borrow. If a household borrows from the home lender, then the profit on a project, if successful at Stage 7, is $y - R_{uh}$ and $y - R_{kh}(q)$ if the household is, respectively, unrevealed and revealed, and 0 if the project is unsuccessful.[87] If borrowing from another lender, these expressions are $y - \tau - R_{ui}$ and $y - \tau - R_{ki}(q)$. The Stage 6 optimal choice is therefore for the unrevealed to borrow from their home lender if $y - R_{uh} > y - \tau - R_{ui}$ and from another

[86] As discussed in footnote 34, we apply a breakeven preference in favor of the home lender to ensure that ε can be taken to 0 in equilibrium.

[87] To save on notations, we drop the limited liability max operators from $\max\{y - R_{uh}, 0\}$ and $\max\{y - R_{kh}(q), 0\}$ and going foward throughout this Stage of the proof: at Stage 5, lenders optimally set rates $\leq y$ as rates above y are certain not to be repaid.

lender otherwise, and for the revealed this condition is $y - R_{kh}(q) > y - \tau - R_{ki}(q)$.

Stage 5. Lenders' breakeven loan rates are $\frac{c}{q}$ for revealed households and $\frac{c}{E[q|u]}$ for the unrevealed.[88] 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 access to household data.[89] 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 $(y - R_{uh}) - (y - \tau - R_{ui}) = -\varepsilon$.[90]

For revealed households, we 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 previous paragraph's limit-pricing argument, the home lender sets R_{kh} according to (8) and all revealed households choose to borrow from it rather than borrowing from other lenders; 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 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

[88] That these are breakeven rates follows from $y > 2c + \tau$ and $q \in \left[\frac{1}{2}, 1\right]$, which imply that $\inf\left(y - \tau - \frac{c}{q}\right) = 0$ and $\inf\left(y - \tau - \frac{c}{E[q|u]}\right) = 0$. Strictly, these are *expected* breakeven rates, given the uncertain realizations of project success. However, with a continuum of borrowers, this distinction is immaterial.

[89] I.e., other lenders are never data access monopolists at Stage 5. See Stage 4 of the proof.

[90] As $\inf\left(y - \tau - \frac{c}{E[q|u]}\right) = 2c - \frac{c}{E[q|u]}$ from $y > 2c + \tau$, and $E[q|u] \geq \frac{1}{2} \Rightarrow \inf\left(2c - \frac{c}{E[q|u]}\right) = 0$.

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 profit on their loan portfolio and therefore will not purchase the data at a positive price.[91] If the DC issuer charges the other lenders a positive fee for data access, then given that the other lenders will not buy the data, the home lender is not 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.[92] 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 the home lender makes on all revealed households in equilibrium is $mE\left[q|\,dc\right]\tau$, where m is the mass of revealed borrowers and $E\left[q|\,dc\right]$ is the expected quality of DC users (and since all revealed households are randomly drawn DC users, this equals the expected quality of revealed households). Both m and $E\left[q|\,dc\right]$ are solved in closed form at Stage 1.

The validity of the conjectured equilibrium therefore centers on the deviation case: will the home lender deviate from the equilibrium by not purchasing the data? If it does, 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 $R_{dev}=$

[91] With $N\geq 3$ and symmetric pure strategies there are always at least two equally informed other lenders engaging in Bertrand competition.

[92] To be precise, it is the full profit minus ε with $\varepsilon\rightarrow 0$.

$R_{uh} = \frac{c}{E[q|u]} + \tau$ the home lender would fail to attract any revealed borrowers, because all revealed households have a better credit quality than the average unrevealed household, as we next show. From (7), we can obtain an expression for the q of the lowest quality revealed household. Only DC users can become revealed and, thus, the lowest quality a revealed household can have is that of the lowest quality DC user. The credit quality of this household is obtained by setting $\varphi = 0$ in (7) because, as can be seen from Figure 2, the DC user with the lowest privacy costs is also the DC user with the lowest credit quality. This yields $q\left(\left(\frac{c}{E[q|u]} + \tau\right) - \tau\right) - c = 0 \Leftrightarrow q = E[q|\, u]$. Hence, the lowest quality revealed household has $q = E[q|\, u]$ and all other revealed households have $q > E[q|\, u]$.[93]

Thus, if the home lender deviates, it must do so with $R_{dev} < \frac{c}{E[q|u]} + \tau$, since it cannot increase its profits by not purchasing the data while charging $R_{dev} = \frac{c}{E[q|u]} + \tau$, as no revealed household would then borrow from it. The deviation, if any, must be in the direction where the home lender accepts lower returns on unrevealed households (by lowering R_{dev}), while gaining new borrowers from the pool of revealed households. However, given $\tau < \frac{1}{3}c$ from (5), $R_{dev} < \frac{c}{E[q|u]} + \tau$ is shown below to be suboptimal for the home lender, as compared to following the outlined equilibrium strategy.[94,95] Using the expression for the profit of the

[93] 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, $y - \tau - \frac{c}{q} > y - \left(\frac{c}{E[q|u]} + \tau\right) \Leftrightarrow q > E[q|\, u]$, which is true for the full mass of revealed households, m (since the revealed household with $q = E[q|\, u]$ has zero weight).

[94] In particular, $\tau < \frac{1}{3}c$ is found as a sufficient condition to ensure this. A necessary and sufficient condition could be found by solving the first order condition of (15) to R_{dev} and finding the highest profit that the deviating lender could make. 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 in (28) for τ small enough, but beyond a point would decline as τ increases). However, the first order condition of (15) to R_{dev} is fourth order in R_{dev} and does not readily lend itself to analysis.

[95] An economic intuition for why an upper bound on τ makes deviation suboptimal is that τ compensates for the adverse selection problem that the uninformed home lender faces when all other lenders buy the data. For example, when the $q = 1$ household borrows from informed other lenders, this household earns $y - \tau - c$ on its successful project. When borrowing from the deviating home lender, it earns $y - R_{dev}$. It therefore 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 the $q = 1$ household, 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 $\frac{c}{E[q|u]} + \tau$. When τ is sufficiently small, this adverse selection problem is severe enough that the home lender does not choose to deviate.

deviating home lender, Π_{dev}, in (15), we find that lost revenues on unrevealed households from cutting R_{dev} always outweigh the revenues on the additional revealed borrowers that choose the home lender when it cuts R_{dev}. Here,

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

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 $q = \frac{c}{R_{dev}-\tau}$.[96] Moreover, $f(q) = \frac{c}{\alpha}\left(\frac{q}{E[q|u]} - 1\right)$.[97] 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 $(1-m)\left(E(q|u)R_{dev} - c\right)$. A sufficient condition for non-deviation is $\frac{\partial \Pi_{dev}}{\partial R_{dev}} > 0$ for $R_{dev} \leq \frac{c}{E[q|u]} + \tau$, because this implies that $R^*_{dev} = \frac{c}{E[q|u]} + \tau$, which equals the equilibrium loan rate. Footnote 98 shows that $\tau < \frac{1}{3}c \Rightarrow \frac{\partial \Pi_{dev}}{\partial R_{dev}} > 0$.[98]

[96] This 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: $y - R_{dev} \geq y - \tau - \frac{c}{q}$.

[97] This comes from the indifference condition that determines which households opt for DC use, as derived in (19). Even though Stage 2 occurs before Stage 4, the same sorting condition applies here, because we are investigating a deviation from the conjectured equilibrium. I.e., backward induction proceeds as if the home lender chooses not to deviate (as below) and then we check (here) that this non-deviation is indeed optimal.

[98] A sufficient condition for $\frac{\partial \Pi_{dev}}{\partial R_{dev}} > 0$ is $\frac{\partial}{\partial R_{dev}}\left[\int_{E(q|u)}^{\frac{c}{R_{dev}-\tau}} 2(qR_{dev} - c)dq + (1-m)\left(\frac{1}{2}R_{dev} - c\right)\right] > 0$, because, firstly, 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 where the indifference frontier as displayed in Figure 2 has "hit" the upper bound of $\varphi = 1$ and the household distribution is uniform). I.e., 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 $f(q) = 2$. Secondly, from (15), $\frac{\partial}{\partial R_{dev}}(1-m)\left(E(q|u)R_{dev} - c\right)$ is always positive and for this term it is therefore sufficient to consider $E(q|u) = \inf E(q|u) = \frac{1}{2}$. Simplifying the sufficient condition in the first sentence of this footnote yields $\frac{R_h - 3\tau}{(R_h-\tau)^3} > \frac{2[E(q|u)]^2 + m - 1}{2c^2}$. This condition is at its tightest when $E(q|u) = \sup E(q|u) = E(q) = \frac{3}{4}$ and $m = \sup m = 1$, so that $\frac{R_{dev} - 3\tau}{(R_{dev}-\tau)^3} > \frac{9}{32c^2}$ is sufficient. The infimum of $\frac{R_{dev} - 3\tau}{(R_{dev}-\tau)^3}$ lies in its corners, because it is a concave function. Investigating these corners, $R_{dev} = \frac{4}{3}c$ and $R_{dev} = 2c + \tau$, 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 $\tau < \frac{c}{3}$.

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

$$q\left(y-\tau I-R_{dc}\left(q\right)\right)-\alpha\varphi\theta>q\max\left(y-\tau I-R_{cash}\right) \tag{16}$$

where $R_{dc}\left(q\right)$ and R_{cash} are, respectively, the loan rates that the household would expect if it chooses DC or cash. The loan rate on cash is not type dependent because cash users do not become revealed and therefore borrow at equilibrium loan rate R_u from (9). Instead, $R_{dc}\left(q\right)=\theta R_k\left(q\right)+\left(1-\theta\right)R_u$, because a DC user's credit quality is revealed with probability θ in which case that household borrows at the equilibrium loan rate $R_k\left(q\right)$ from (8), while with probability $1-\theta$ the household remains unrevealed and pays R_u on its loan. We can now simplify (16) to:

$$\alpha\varphi<q\left(R_u-R_k\left(q\right)\right) \tag{17}$$

and replacing from equations (8) and (9), this becomes the sorting condition in (7).

Stage 1. At Stage 1, the DC issuer optimally sets $mE\left[q|\, dc\right]\tau$ as access fee to the home lender and provides the data for free to the other lenders, as follows from the Stage 4 derivations.[99] The DC issuer also determines θ at Stage 1 by maximizing $mE\left[q|\, dc\right]\tau$ to θ.

We here derive closed form solutions for m and for $E\left[q|\, dc\right]$ and show that $\frac{\partial mE\left[q|dc\right]\tau}{\partial\theta}>0$ and therefore $\theta=1$ is optimal. We first write m as a function of θ and note that $m\left(\theta\right)=\theta\mu\left(\theta\right)$ where $\mu\left(\theta\right)$ is the mass of DC users, which we can find by integrating over the indifference frontier displayed as the unbroken line in Figure 2. 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

[99] Namely, 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 $mE\left[q|\, dc\right]\tau-\varepsilon$ with $\varepsilon\rightarrow0$.

of φ) are DC users. From the first to the second point, we integrate over the indifference frontier and from the second point to $q = 1$ we integrate over a uniformly distributed mass of households. The first point has been established at Stage 4 and is given by $q = E[q|\, u]$. The second point is derived from entering $\varphi = 2$ in (7) and setting (7) to equality, which yields $2\alpha = c\left(\frac{q}{E[q|u]} - 1\right)$ and can be written to $q = \frac{2\alpha+c}{c} E[q|\, u]$. 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).[100] 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

$$\mu(\theta) = \int_{E[q|u]}^{\frac{2\alpha+c}{c} E[q|u]} f(q)\, dq + \int_{\frac{2\alpha+c}{c} E[q|u]}^{1} 2dq \tag{18}$$

$$f(q) = \frac{c}{\alpha}\left(\frac{q}{E[q|\, u]} - 1\right) \tag{19}$$

where $f(q)$ is derived from setting (7) to equality and writing to φ. Moreover, the 2 in the second integral in (18) comes from the uniform distribution over $q \in \left[\frac{1}{2}, 1\right]$. We note that $E[q|\, u]$ is a function of θ (derived further below). Solving (18) gives:

$$\mu(\theta) = 2\left(1 - \frac{c+\alpha}{c} E[q|\, u]\right) \tag{20}$$

and thus

$$m(\theta) = 2\theta\left(1 - \frac{c+\alpha}{c} E[q|\, u]\right) \tag{21}$$

where the term in parentheses 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 come to a closed-form expression for $E[q|\, u]$, we first note that $(1 - m(\theta))\, E[q|\, u] + m(\theta)\, E[q|\, dc] = E[q] = \frac{3}{4}$, which can be written to

$$E[q|\, u] = \frac{3/4 - m(\theta)\, E[q|\, dc]}{1 - m(\theta)} \tag{22}$$

[100] 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).

where $E\left[q|\, dc\right]$ is the expected credit quality of DC users, which can be found as a weighted average of expected values with respect to the integral areas identified in (18):

$$E\left[q|\, dc\right]=\frac{\int\limits_{E[q|u]}^{\frac{2\alpha+c}{c}E[q|u]} q\left(\frac{c}{\alpha}\left(\frac{q}{E[q|u]}-1\right)\right)dq+\int\limits_{\frac{2\alpha+c}{c}E[q|u]}^{1} 2qdq}{\int\limits_{E[q|u]}^{\frac{2\alpha+c}{c}E[q|u]} \frac{c}{\alpha}\left(\frac{q}{E[q|u]}-1\right)dq+\int\limits_{\frac{2\alpha+c}{c}E[q|u]}^{1} 2dq} \tag{23}$$

and solving this gives

$$E\left[q|\, dc\right]=\frac{4\alpha^2\left(E\left[q|\, u\right]\right)^2+3c^2\left(\left(E\left[q|\, u\right]\right)^2-1\right)+6\alpha c\left(E\left[q|\, u\right]\right)^2}{6c\left(\alpha E\left[q|\, u\right]+c\left(E\left[q|\, u\right]-1\right)\right)} \tag{24}$$

Replacing from (21) and (24) into (22) and solving leads to:

$$E\left[q|\, u\right]=\frac{16\alpha^2\theta\left(E\left[q|\, u\right]\right)^2+3c^2\left(3+4\theta\left(\left(E\left[q|\, u\right]\right)^2-1\right)\right)+24\alpha c\theta\left(E\left[q|\, u\right]\right)^2}{12c\left(2\alpha\theta E\left[q|\, u\right]+2c\theta\left(E\left[q|\, u\right]-1\right)+c\right)} \tag{25}$$

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

$$E\left[q|\, u\right]=\frac{c}{\theta}\frac{3c\left(2\theta-1\right)+\sqrt{3}\sqrt{3c^2\left(1-\theta\right)+4\alpha^2\theta\left(4\theta-3\right)}}{6c^2+8\alpha^2} \tag{26}$$

where an additional negative root solution can be discarded as it always gives $E\left[q|\, u\right]<\frac{1}{2}$, which violates $q\in\left[\frac{1}{2},1\right]$.[101] It can also be verified that for the solution in (26), it holds that $E\left[q|\, u\right]\in\left[\frac{1}{2},\frac{3}{4}\right]$.[102] We also note that the apparent singularity at $\theta=0$ is not a genuine singularity. Its appearance comes from the fact that (26) is the solution of (25) when (25) is written as a quadratic equation. However, at $\theta=0$, (25) 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 necessarily unrevealed

[101] This can be seen from the fact that the negative root solution (replacing the first plus in the numerator of (26) with a minus) is smaller than $\frac{2\theta-1}{\theta}\frac{3c^2}{6c^2+8\alpha^2}$ and $\frac{2\theta-1}{\theta}\frac{3c^2}{6c^2+8\alpha^2}<\frac{2\theta-1}{\theta}\frac{3c^2}{6c^2}=1-\frac{1}{2}\frac{1}{\theta}<\frac{1}{2}$.

[102] Using the $MaxValue$ and $MinValue$ functions in *Mathematica* on (26) 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.

and $E[q|u] = E[q] = \frac{3}{4}$. Furthermore, we note that $\frac{\partial E[q|u]}{\partial\theta} < 0$ and $\frac{\partial^2 E[q|u]}{\partial\theta^2} < 0$.[103] Next, we replace from (26) into (21) to obtain

$$m(\theta) = \frac{3c(c + 3\alpha(1-2\theta)) - 8\alpha^2\theta - (\alpha + c)\sqrt{3}\sqrt{3c^2(1-\theta) + 4\alpha^2\theta(4\theta - 3)}}{3c^2 - 4\alpha^2} \tag{27}$$

where we verify that $m(\theta) \in [0, 1]$ over our parameter space.[104]

This provides us with the elements needed to show that $\frac{\partial m(\theta)E[q|dc]\tau}{\partial\theta} > 0$ and therefore $\theta = 1$ is optimal. In particular, replacing for $m(\theta)$ from (27) and for $E[q|dc]$ from (24) and (26), and simplifying, we obtain the closed form for the optimal data selling fee of the DC issuer to the home lender, denoted by $\widehat{\Omega}$:

$$\widehat{\Omega} = \frac{\tau}{4\theta(3c^2 - 4\alpha^2)^2}\left(\begin{array}{c} 9c^4(5\theta - 2) - 18\alpha c^3(2 + \theta(4\theta - 5)) + 48\alpha^4\theta \\ +24\alpha^3 c\theta(3 - 4\theta) + 24\alpha^2c^2\left(1 - 4\theta + 8\theta^2\right) + \\ c(8\alpha^2 + 6c^2 + 12\alpha c)(1 - 2\theta)\sqrt{9c^2(1-\theta) + 12\alpha^2\theta(4\theta - 3)} \end{array}\right) \tag{28}$$

and using (28) we can confirm that $\frac{\partial\widehat{\Omega}}{\partial\theta} > 0$.[105]

It is useful to also record $R_u(\theta)$ (from replacing (26) into (9)) and its derivatives.[106,107]

$$R_u(\theta) = \theta\frac{6c^2 + 8\alpha^2}{3c(2\theta - 1) + \sqrt{3}\sqrt{3c^2(1-\theta) + 4\alpha^2\theta(4\theta - 3)}} + \tau \tag{29}$$

$$\frac{\partial R_u(\theta)}{\partial\theta} > 0; \frac{\partial^2 R_u(\theta)}{\partial\theta^2} > 0; \frac{\partial R_u(\theta)}{\partial\alpha} < 0; \frac{\partial R_u(\theta)}{\partial c} > 1; \frac{\partial R_u(\theta)}{\partial\tau} = 1 \tag{30}$$

[103] Using the $MaxValue$ function in *Mathematica*, we verify that $\sup_{\theta\in(0,1],c\geq 1,\alpha\in\left[0,\frac{c}{6}\right]} \frac{\partial E[q|u]}{\partial\theta} < 0$ and $\sup_{\theta\in(0,1],c\geq 1,\alpha\in\left[0,\frac{c}{6}\right]} \frac{\partial^2 E[q|u]}{\partial\theta^2} < 0$.

[104] Using the $MinValue$ and $MaxValue$ functions in *Mathematica*, we find the infimum and supremum of (27) as $\inf_{\theta\in[0,1],c\geq 1,\alpha\in\left[0,\frac{c}{6}\right]} m(\theta) = 0$ and $\sup_{\theta\in[0,1],c\geq 1,\alpha\in\left[0,\frac{c}{6}\right]} m(\theta) = 1$.

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

[106] 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 $\theta \in [0, 1], c \geq 1, \alpha \in \left[0, \frac{c}{6}\right]$.

[107] For empirical estimates on the relation between loan rates and the extent of information asymmetry (inversely related to θ in our framework), see DeFusco et al. (2022).

and footnote 108 discusses the intuitions behind these comparative statics.[108] ■

Proof of Proposition 2. When $y < \frac{4}{3}c$, there exists no breakeven R_u and lenders are unwilling to lend to unrevealed households.[109] This simplifies the problem as compared to Proposition 1.

Stage 6. If a revealed household borrows from its home lender or from another lender 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 is therefore to borrow from the home lender if $y - R_{kh}(q) > y - \tau - R_{ki}(q)$ and from another lender otherwise.

Stage 5. We can distinguish three cases according to the creditworthiness of a household. First, if $q < \frac{c}{y}$, then a household's project has negative NPV ($qy < c$) and no lender will lend to it. Second, if $q \in \left[\frac{c}{y}, \frac{c}{y-\tau}\right]$, then the household's project only has positive NPV when borrowing on-island and therefore only the home lender will make it a loan rate offer. Third, if $q > \frac{c}{y-\tau}$ then all lenders can make loan offers: with breakeven rate $\frac{c}{q}$ the project's expected net return, $q\left(y - \tau - \frac{c}{q}\right)$, is positive. For households with $q > \frac{c}{y-\tau}$, then, the same arguments as in the Proof of Proposition 1 apply to this stage and therefore the equilibrium loan rate is given by (8), which 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). We note that $q > \frac{c}{y-\tau}$ households exist if and only if $\tau < y - c$ (given $q \leq 1$).

Stage 4. The deviation case where the home lender refuses to purchase the data can be

[108] A higher θ worsens the expected quality of the pool of the unrevealed, implying a higher $R_u(\theta)$, and cascade effects become stronger as θ rises further, leading to convexity; 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 funding costs, c, translate into higher loan rates; and, lastly, greater home lender market power, τ, translates into higher loan rates as well.

[109] 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 $\inf \frac{c}{E[q|u]} = \frac{4}{3}c > y$.

directly excluded. If an uninformed home lender sets $R_{dev} < y$ to attract revealed households, this immediately induces the full set of unrevealed households to borrow from the home lender too, including many bad borrowers. This implies that 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 $R_{dev} < y < \frac{4}{3} c$. The same arguments as in the Proof of Proposition 1 then 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, the general expression for this fee is $m(\theta)\, E[q|\, dc]\, \tau$, but its closed form is different than in Proposition 1 and is solved in (36).

Stage 2. Households with $q < \frac{c}{y}$ choose cash as they cannot obtain loans at Stage 5 and thus see only a privacy cost to using DC. The same applies to households with $q \in \left[\frac{c}{y}, \frac{c}{y-\tau}\right]$, because if they use DC and become revealed their loan rate will be $R_{kh}(q) = y$, which leaves no profit for them from a successful project. If $\tau > y - c$ and 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 $\tau < y - c$, then at Stage 2 households with $q > \frac{c}{y-\tau}$ prefer DC over cash if and only if $q(y - \tau I - R_{dc}(q)) - \alpha\varphi\theta > 0$, which can be written to $\theta q \left(y - \left(\frac{c}{q} + \tau\right)\right) - \alpha\varphi\theta > 0$ and from there to

$$\alpha\varphi < q(y - \tau) - c \tag{31}$$

where we note that (31) can alternatively be written as $q > \frac{\alpha\varphi + c}{y-\tau}$.[110]

Stage 1. As in the Proof of Proposition 1, we here 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

[110] Since $\alpha\varphi \geq 0$, we have that $\frac{\alpha\varphi+c}{y-\tau} \geq \frac{c}{y-\tau}$ and therefore the condition in (31) 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.

show that $\frac{\partial\widehat{\Omega}}{\partial\theta} > 0$, from which $\theta = 1$ 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 to use the DC or not. From (31), we have that this household chooses DC if $y > 2\alpha + c + \tau$ and chooses cash if $y < 2\alpha + c + \tau$.[111]

If $y < 2\alpha + c + \tau$, then even for $q = 1$ there are still households who prefer cash over DC. In this case, $m(\theta)$ and $E[q|\, dc]$ are given by, respectively, (32) and (33).[112]

$$m(\theta) = \theta \int_{\frac{c}{y-\tau}}^{1} \left(\frac{q(y-\tau)-c}{\alpha}\right) dq = 2\theta\left(1 - \frac{\alpha+c}{y-\tau}\right) \tag{32}$$

$$E[q|\, dc] = \int_{\frac{c}{y-\tau}}^{1} q\left(\frac{q(y-\tau)-c}{\alpha}\right) dq = \frac{(y-c-\tau)^2\,(2(y-\tau)+c)}{6\alpha\,(y-\tau)^2} \tag{33}$$

Instead, when $y > 2\alpha + c + \tau$, we have:

$$m(\theta) = \theta\left(\int_{\frac{c}{y-\tau}}^{\frac{2\alpha+c}{y-\tau}} \left(\frac{q(y-\tau)-c}{\alpha}\right) dq + \int_{\frac{2\alpha+c}{y-\tau}}^{1} 2dq\right) = \theta\left(\frac{(y-c-\tau)^2}{2\alpha\,(y-\tau)}\right) \tag{34}$$

$$E[q|\, dc] = \frac{\int_{\frac{c}{y-\tau}}^{\frac{2\alpha+c}{y-\tau}} q\left(\frac{q(y-\tau)-c}{\alpha}\right) dq + \int_{\frac{2\alpha+c}{y-\tau}}^{1} 2qdq}{\int_{\frac{c}{y-\tau}}^{\frac{2\alpha+c}{y-\tau}} \frac{q(y-\tau)-c}{\alpha} dq + \int_{\frac{2\alpha+c}{y-\tau}}^{1} 2dq} = \frac{3(y-\tau)^2 - 4\alpha^2 - 6\alpha c - 3c^2}{6(y-\tau)(y-\alpha-c-\tau)} \tag{35}$$

Taking together (32) and (33) simplifying, and idem for (34) and (35), this means that:

$$\widehat{\Omega} = m(\theta)\, E[q|\, dc]\,\tau = \begin{cases} \theta\tau\frac{\left(3(y-\tau)^2-4\alpha^2-6\alpha c-3c^2\right)}{3(y-\tau)^2} & \text{if } y < 2\alpha+c+\tau \\ \theta\tau\frac{(y-c-\tau)^4(2(y-\tau)+c)}{12\alpha^2(y-\tau)^3} & \text{if } y > 2\alpha+c+\tau \end{cases} \tag{36}$$

In either case, $\widehat{\Omega}$ increases linearly as θ rises, and $\theta = 1$ is 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 $\int_{\frac{1}{2}}^{1} 2(qy_h - c)\, dq = \frac{3}{4}y_h - c$, because all households are borrowers in Proposition 1.[113]

[111] We note from (5) and (6) that $\tau \in \left[0, \frac{1}{3}c\right]$ and $\alpha \in \left[0, \frac{1}{6}c\right]$, 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.

[112] Here, $\left(\frac{q(y-\tau)-c}{\alpha}\right)$ comes from setting (31) to equality and solving to φ.

[113] I.e., $f(q|\,\text{"borrowing"}) = 2$ in (11). Also note that $\frac{3}{4}y_h - c$ is strictly positive given $y_h > 2c + \tau$.

The second term in (11) follows from rewriting the indifference frontier (wherein (7) holds with equality) to integrate over $\varphi \in [0,2]$, which is $1 - \frac{\alpha\varphi+c}{c}E[q|\,u]$.[114] Hence, the second term in (11) becomes $\int_0^2 (\alpha\varphi\theta)\left(1 - \frac{\alpha\varphi+c}{c}E[q|\,u]\right)d\varphi = 2\alpha\theta\left[1 - \left(1+\frac{4\alpha}{3c}\right)E[q|\,u]\right]$, which we note is always a positive term given $E[q|\,u] \leq \frac{3}{4}$ and $\alpha < \frac{1}{6}c$.[115] Putting terms together:

$$W_h = \frac{3}{4}y_h - c - 2\alpha\theta\left[1 - \left(1 + \frac{4\alpha}{3c}\right)E[q|\,u]\right] \tag{37}$$

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 (32) and (34), and therefore the first term in (11) is $\theta\int_{\frac{c}{y_l-\tau}}^1 (qy_l - c)\left(\frac{q(y_l-\tau)-c}{\alpha}\right)dq$ if $y_l < 2\alpha + c + \tau$; and $\theta\left(\int_{\frac{c}{y_l-\tau}}^{\frac{2\alpha+c}{y_l-\tau}} (qy_l - c)\left(\frac{q(y_l-\tau)-c}{\alpha}\right)dq + \int_{\frac{2\alpha+c}{y_l-\tau}}^1 2(qy_l-c)\,dq\right)$ if $y_l > 2\alpha + c + \tau$, which become:

$$W_l\ (\text{first term}) = \begin{cases} \theta\frac{(y_l-c-\tau)^2(2y_l(y_l-c-\tau)+3\tau c)}{6\alpha(y_l-\tau)^2} & \text{if } y_l < 2\alpha + c + \tau \\ \theta\left(y_l\left(1 + \frac{4\alpha^2}{3(y_l-\tau)^2}\right) + \frac{c^2(y_l-2\tau)}{(y_l-\tau)^2} - 2c\right) & \text{if } y_l > 2\alpha + c + \tau \end{cases} \tag{38}$$

For W_l, for the second term in (11), we have that (32) and (34) represent the integrals for DC users as defined for integration over q, which we rewrite for integration over φ, with the density function taken from setting (31) to equality.[116] This gives $\int_0^{\frac{y_l-\tau-c}{\alpha}} (\alpha\varphi\theta)\left(1 - \frac{\alpha\varphi+c}{y_l-\tau}\right)d\varphi$ if $y_l < 2\alpha + c + \tau$; and $\int_0^2 (\alpha\varphi\theta)\left(1 - \frac{\alpha\varphi+c}{y_l-\tau}\right)d\varphi$ if $y_l > 2\alpha + c + \tau$, which become:

$$W_l\ (\text{second term}) = \begin{cases} \theta\frac{1}{6\alpha}\left(\frac{(y_l-\tau-c)^3}{y_l-\tau}\right) & \text{if } y_l < 2\alpha + c + \tau \\ \theta\frac{2\alpha}{3}\left(\frac{3y_l-3\tau-4\alpha-3c}{y_l-\tau}\right) & \text{if } y_l > 2\alpha + c + \tau \end{cases} \tag{39}$$

[114] 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 $\frac{\alpha\varphi+c}{c}E[q|\,u] - \frac{1}{2}$ below the indifference frontier and since $q \in \left[\frac{1}{2}, 1\right]$ there is a mass of $\frac{1}{2} - \left(\frac{\alpha\varphi+c}{c}E[q|\,u] - \frac{1}{2}\right)$ above it, which are DC users. The latter simplifies to $1 - \frac{\alpha\varphi+c}{c}E[q|\,u]$.

[115] As these imply $\sup\left(1 + \frac{4\alpha}{3c}\right)E[q|\,u] = \frac{11}{12} < 1$.

[116] 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.

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

$$\lambda = \begin{cases} \frac{1}{6\alpha}\left(\frac{(y_l-\tau-c)^2(y_l(y_l-c)+\tau(2c-\tau))}{y_l-\tau}\right) & \text{if } y_l < 2\alpha+c+\tau \\ y_l + \frac{y_l-2\tau}{(y_l-\tau)^2}\left(c^2+2\alpha c+\frac{4}{3}\alpha^2\right) - 2(\alpha+c) & \text{if } y_l > 2\alpha+c+\tau \end{cases} \tag{40}$$

where we note that λ is always positive here.[117] ■

Proof of Proposition 5. First, W_h in (13) is necessarily maximized at $\theta = 0$, as this sets the negative second term (privacy costs) to 0.[118] Hence, $\gamma = 1 \Rightarrow \theta^* = 0$. Second, it follows directly 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).[119]

$$\theta^* = \frac{3}{32\alpha^2}\left[4\alpha^2 + c^2 - \sqrt{\frac{(c^2-12\alpha^2)\left(8\alpha^2\gamma(\alpha+c)+(3c^2-4\alpha^2)(1-\gamma)\lambda\right)^2}{16\alpha^4\gamma^2+48\alpha^2(\alpha+c)(1-\gamma)\gamma\lambda+3(3c^2-4\alpha^2)(1-\gamma)^2\lambda^2}}\right] \tag{41}$$

■

[117] In particular, $\frac{(y_l-\tau-c)^2(y(y-c)+\tau(2c-\tau))}{y_l-\tau}$ is immediately positive, while $\inf\left\{y_l + \frac{y_l-2\tau}{(y_l-\tau)^2}\left(c^2+2\alpha c+\frac{4}{3}\alpha^2\right) - 2(\alpha+c)\right\}$ given $y_l > 2\alpha+c+\tau$ occurs at $y_l \to 2\alpha+c+\tau$ where $\lambda = \frac{2\alpha(4\alpha(\alpha+\tau)+c(2\alpha+3\tau))}{(2\alpha+c)^2} > 0$.

[118] This can also be seen by replacing $R_u(\theta)$ from (9) and rewriting the last term in (37) 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. These 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 $1-\frac{3}{4}-\frac{3}{4}\frac{1}{6}\frac{3}{4} > 0$.

[119] Here we continue using the term λ as a collection of constants given by (40). Also note that a second, positive root solution can be excluded because it implies $\theta^* > 1$.

B Extensions

B.1 Alternative sources of credit quality data

In the baseline model, the DC is the unique source of credit quality information. We here consider two alternative sources: pre-existing public information and public information that is gathered through an alternative payment system technology.[120]

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 3.5. I.e., 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 (provided these have positive NPV projects) 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: for the set of households that is not "pre-revealed", the model is identical to the baseline. Lender and DC issuer total profits are affected, but only because the pool of households that is not pre-revealed is smaller. With a renormalization, total profits remain the same as in the baseline too: 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 payment technology that creates public information about credit quality. Household payment choice then centers on DC in comparison to this alternative means of payment. This could, for example, be a data-sharing CBDC (as discussed in Sections 2 and 8) or deposits under open banking policies (as further considered in Appendix B.6). 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 focus on the case where $\theta_{alt} < 1$ and $\theta_{dc} > \theta_{alt}$: the DC is more

[120] We restrict attention to public information. Lender-specific data on a subset of households (such as depositors) would significantly complicate the model and lies outside the scope of the current analysis.

informationally intrusive than the alternative.

Here, a household chooses DC over the alternative if $q\left(y-\tau I-\left(\theta_{dc}R_k\left(q\right)+\left(1-\theta_{dc}\right)R_u\right)\right)-\alpha\varphi\theta_{dc}>q\left(y-\tau I-\left(\theta_{alt}R_k\left(q\right)+\left(1-\theta_{alt}\right)R_u\right)\right)-\alpha\varphi\theta_{alt}$, which simplifies to $\alpha\varphi<q\left(R_u-R_k\left(q\right)\right)$ and this is identical to (17). That is, the condition that previously identified household payment choice between DC and cash, now identifies household payment choice between DC and the alternative technology. Similarly, following through on other steps in Appendix A, nothing changes about the DC issuer optimization. For instance, (28) sees θ replaced with $\theta_{dc}-\theta_{alt}$ and therefore $\theta_{dc}=1$ remains privately optimal.[121]

B.2 Correlation of credit quality and privacy preferences

In the baseline model, the two dimensions of household heterogeneity, q and φ, are assumed to be independent of each other. This extension investigates correlation between household credit quality and privacy preferences.[122] In the interest of tractability, we center attention on the cases of a perfectly positive and perfectly negative correlation of these features. In either case, the two dimensions of household heterogeneity collapse to a single dimension. In particular, $\varphi=4q-2$ represents a perfectly positive correlation, where $q=\frac{1}{2}$ implies $\varphi=0$, while $q=1$ implies $\varphi=2$. Conversely, $\varphi=4\left(1-q\right)$ represents a perfectly negative correlation, where $q=\frac{1}{2}$ implies $\varphi=2$, while $q=1$ implies $\varphi=0$.[123]

With a perfectly negative correlation the model collapses to a fully unraveling Lemons' Market.[124] With a single dimension of households it is more challenging to generate a mechanism that "halts" cascades. This can only happen if high credit quality households, which have incentives to type differentiate, also care more about privacy, so that some of them sort into cash use, thereby cross-subsidizing lower quality households that choose to

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

[122] For empirical studies that explore correlation between measures of household privacy and quality, see Chen et al. (2021) and Lin (2022).

[123] We report here on the main outcomes of the derivations, which are performed in a *Mathematica* file that is available on request.

[124] The indifference condition for choosing DC over cash becomes $q\geq\frac{E[q|u](4\alpha+c)}{c+4\alpha E[q|u]}$ and solving for $E\left[q|\,u\right]$ now leads to outcomes where $E\left[q|\,u\right]>1$: there is no $E\left[q|\,u\right]\in\left[\frac{1}{2},\frac{3}{4}\right]$ that solves the fixed point problem, because the Lemons' Market fully unravels.

use cash. That is, when quality and care for privacy are positively correlated. In the case of a perfectly positive correlation, the indifference condition for choosing DC over cash becomes $q \geq \frac{E[q|u](c-2\alpha)}{c-4\alpha E[q|u]}$ and solving for $E[q|u]$ using the same steps as before gives $E[q|u] = \frac{\sqrt{c}\sqrt{c+16\alpha}-c}{8\alpha}$.[125] Here, $\theta = 1$ remains privately optimal for the DC issuer.[126] Recalculating welfare, W_h and W_l, respectively, decline and rise as θ increases, like in the baseline. However, unlike the baseline model W_h is now linear in θ (as is W_l). This implies that θ^* is either 0 or 1.

B.3 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

$$u(q, \varphi) = q \max\{(y - \tau I - R), 0\} - \alpha\varphi\Theta^2 \tag{42}$$

As the DC becomes more intrusive, households experience a more than proportional increase in their privacy costs. For Proposition 1, nothing changes in Stages 3 - 7 of the game depicted in Section 3.5. In Stage 2, the derived indifference frontier in (7) now becomes

$$\theta\varphi \leq \frac{c}{\alpha}\left(\frac{q}{E[q|u]} - 1\right) \tag{43}$$

where for $\theta = 1$ we have that (7) and (43) 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 $\theta = 1$ indeed remains optimal for the DC

[125] Depending on parameter values, this can imply either full or partial cascades towards DC use. In particular, when α is small, full unraveling towards DC use occurs. Instead, when α is large enough, the privacy preferences of higher quality households are strong enough to counteract the cascades and sustain a degree of cash use. For example, when $c = 1$ and $\alpha = \frac{1}{6}$ (the upper bound of α), households with $q < 0.84$ choose cash and $E[q|u] = 0.69$.

[126] The DC issuer's optimal data access fees, $\widehat{\Omega} = m(\theta) E[q|dc]\tau$, unambiguously increase as θ rises: while $E[q|dc]$ and the mass of DC users, μ, are now constant terms that are unaffected by θ, the mass of revealed borrowers is given by $m(\theta) = \theta\mu$ and therefore increases as θ rises.

issuer.[127]

For Proposition 2, equation (31) now becomes

$$\alpha\varphi\theta < q\left(y-\tau\right)-c \tag{44}$$

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

$$\mu\left(\theta\right)=\begin{cases} 2\left(1-\frac{c+\alpha\theta}{y-\tau}\right) & \text{if } \theta<\frac{y-\tau-c}{2\alpha} \\ \frac{1}{2\alpha\theta}\left(\frac{\left(y-c-\tau\right)^2}{y-\tau}\right) & \text{if } \theta>\frac{y-\tau-c}{2\alpha} \end{cases},\; m\left(\theta\right)=\begin{cases} 2\theta\left(1-\frac{c+\alpha\theta}{y-\tau}\right) & \text{if } \theta<\frac{y-\tau-c}{2\alpha} \\ \frac{1}{2\alpha}\left(\frac{\left(y-c-\tau\right)^2}{y-\tau}\right) & \text{if } \theta>\frac{y-\tau-c}{2\alpha} \end{cases} \tag{45}$$

where we note that for both $\mu\left(\theta\right)$ and $m\left(\theta\right)$ 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 $\mu\left(\theta\right)$ and $m\left(\theta\right)$ respond differently to θ than in the baseline provides additional insight on some of the mechanisms underlying the model. In particular, in the low y case (Proposition 2), we now have that $\frac{\partial\mu(\theta)}{\partial\theta}<0$.[128] In this case, the quadratic form makes marginal privacy costs large enough that more households are repelled from DC use in response to a higher θ 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 have that $\frac{\partial m(\theta)}{\partial\theta}=0$, as the decline in the mass of DC users is exactly offset by the higher revelation probability 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.[129] Overall, $\theta=1$ remains optimal for the DC issuer in the low y case. Even when $\frac{\partial m(\theta)}{\partial\theta}=0$, optimal data access fees $\left(m\left(\theta\right)E\left[q|\,dc\right]\tau\right)$ continue to rise as

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

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

[129] I.e., the decline in the mass of DC users is more than offset by the higher probability of revelation per DC user.

θ increases, because higher q households then remain among the DC users $\frac{\partial E[q|dc]}{\partial \theta} > 0$.

B.4 Fees or subsidies on DC use

The baseline model centers on a data monopoly that extracts information rents. An alternative possibility is that the DC issuer could make profits by charging fees to households for using the DC.[130] Conversely, the DC issuer may also find value in considering subsidies on DC use to increase the mass of revealed households whose data it can market to lenders. Let F represent the fee that the DC issuer charges households, where $F > 0$ represents a positive fee, $F < 0$ represents a subsidy, and $F = 0$ represents the baseline model. Within the timing of the game shown in Section 3.5, the DC issuer sets F at Stage 1 (together with its other decision variables) and F is paid out at Stage 2 when households decide on their means of payment.[131]

The main insight from this addition is that the incentive of the DC issuer to charge a positive fee increases as τ decreases. The lower is τ, the less market power home lenders have and the closer the loan market moves to a perfectly competitive outcome where there are no information rents for the DC issuer to extract. In the limit case of $\tau \to 0$, only fees can make the DC issuer a profit and its optimal fee can be found from maximizing the product of F and $m(\theta, F)$, which is the mass of DC users as a function of θ and F.[132] Instead, when market power in lending is larger (but still small enough to prevent credit market monopolization), the DC issuer's information rents rise, as do its incentives to cross-subsidize information

[130] See also Verdier (2020), who considers DC fee optimization in a setting without information rents.

[131] We abstract here from considerations of the sources of funds (i.e., the funds that the DC issuer can make available towards paying a subsidy or the funds that households have available to pay a fee for DC use).

[132] Following the same steps as before, we get that households choose DC over cash if $\alpha\varphi + \frac{F}{\theta} \leq c\left(\frac{q}{E[q|u]} - 1\right)$ in Proposition 1 and $q \geq \frac{\alpha\varphi + c + F}{y - \tau}$ in Proposition 2, while $\theta = 1$ remains optimal for the DC issuer. From here, closed form solutions for $m(\theta, F)$ can be found following the same steps as in (27), (32) and (34).

acquisition with F.[133]

B.5 Monetary policy

Monetary policy influences bank funding costs and we can therefore consider a policy maker who determines c after the realization of the state. We assume that this policy maker has no other objectives for monetary policy than to maximize aggregate welfare in our model and we allow for arbitrarily negative policy rates, implying $c \in (0, 1)$ is possible.

The outcome is that monetary policy can help prevent the bad state from taking hold but cannot, on its own, overcome the data externalities that dominate welfare in the good state. When the low y materializes, the monetary authority can always set c low enough to prevent Proposition 2's constrained-credit equilibrium from emerging. In particular, given $y > \tau$ (in fact, $y > 3\tau$ by (4) and (5)), a sufficiently small c can be found to bring about $y > 2c + \tau$, and thereby the equilibrium of Proposition 1.[134] With monetary policy leaning against the bad state, disclosure loses it social value, because credit provision is assured and the social benefits of type differentiation vanish. Monetary policy can only influence aggregate credit access, however, and has no ability to prevent disclosure cascades. Instead, a combination of $\theta^* = 0$ (banning the DC) and a state-dependent monetary policy that prevents the equilibrium with limited credit access, attains the first-best.

[133] The first order condition for the optimal fee is $\frac{\partial F m(\theta,F)}{\partial F} + \frac{\partial m(\theta,F) E[q|dc]\tau}{\partial F} = 0$ where the first term is the derivative of DC fee revenue and the second term is the derivative of data access fees. We can write this further to $\left(F\frac{\partial m(\theta,F)}{\partial F} + m(\theta,F)\right) + \tau\left(\frac{\partial m(\theta,F)}{\partial F} E[q|\, dc] + \frac{\partial E[q|dc]}{\partial F} m(\theta,F)\right) = 0$. This derivative can be evaluated at $F = 0$ to answer the question of whether a fee or subsidy will be offered. Moreover, we note that $\frac{\partial m(\theta,F)}{\partial F} < 0$. Overall, this becomes $\frac{1}{\tau} + \frac{\partial E[q|dc]}{\partial F} = \left(-\frac{\partial m(\theta,F)}{\partial F}\right)\frac{E[q|dc]}{m(\theta,F)}$.

[134] Of course, this is a partial reasoning, because it does not account for any costs to cutting policy rates, such as inflationary pressure.

B.6 Portable data

This extension considers an alternative setup where households are given ownership of their data.[135] The purpose of this extension is to highlight the differences between a free data porting market and a data monopoly in our framework. This extension lets each household, k, choose its own probability, θ_k, that its type q becomes revealed to all lenders. Here, there is no choice between payment instruments with various degrees of disclosure. Instead, households themselves decide on their preferred degree of revelation. Consider a payment service provider that always learns q and a household that can tell it with what probability to reveal its data to lenders. In essence, the payment service provider acts as a verification mechanism for households to potentially reveal their private information on their type.[136]

We embed this setting in the quadratic privacy cost case of Appendix B.3, because this enables the derivation of an interior solution for a household's optimal degree of disclosure. This extension has a 4-stage game where in Stage 1 households choose θ_k, while Stages 2-4 on lender loan rates, household borrowing and loan repayment are identical to Stages 5-7 in Section 3.5. A household's optimization problem is:

$$\max_{\theta_k \in [0,1]} \left\{ \theta_k q \max\left\{ \left(y - \tau I - R_k\left(q\right)\right), 0 \right\} + \left(1 - \theta_k\right) q \max\left\{ \left(y - \tau I - R_u\right), 0 \right\} - \alpha \varphi \theta_k^2 \right\} \quad (46)$$

We here focus on the high y case in Table 1, because it is household externalities that are of interest in this extension. In the high y case, the loan market equilibrium continues to be represented by equations (8) and (9). The optimization problem in (46) then becomes

$$\max_{\theta_k \in [0,1]} \left\{ \theta_k q \left(y - \frac{c}{q} - \tau \right) + \left(1 - \theta_k\right) q \left(y - \frac{c}{E\left[q \middle| u\right]} - \tau \right) - \alpha \varphi \theta_k^2 \right\} \quad (47)$$

[135] Data portability policies (allowing households to take their data to other lenders) have been implemented in various countries (OECD, 2023) and have been investigated theoretically and empirically in several recent papers (Babina et al., 2022; Brunnermeier and Payne, 2022; Garratt and Lee, 2021; He et al., 2023; Nam, 2022). Relatedly, see Bank et al. (2023) and Qi et al. (2023) on the impact of introducing credit registries (which, in essence, "port" part of household data) on loan market competition.

[136] This bears resemblance to a role foreseen for the digital euro: "digital euro users would retain control over the use of their data by PSPs, with the possibility of opting for PSP data usage for the provision of additional services" (ECB, 2023).

and is solved as

$$\theta_k = \begin{cases} \min\left\{\frac{1}{2}\left(\frac{c}{\alpha\varphi}\left(\frac{q}{E[q|u]}-1\right)\right),1\right\} & \text{if } q > E\left[q|\, u\right] \\ 0 & \text{if } q < E\left[q|\, u\right] \end{cases} \tag{48}$$

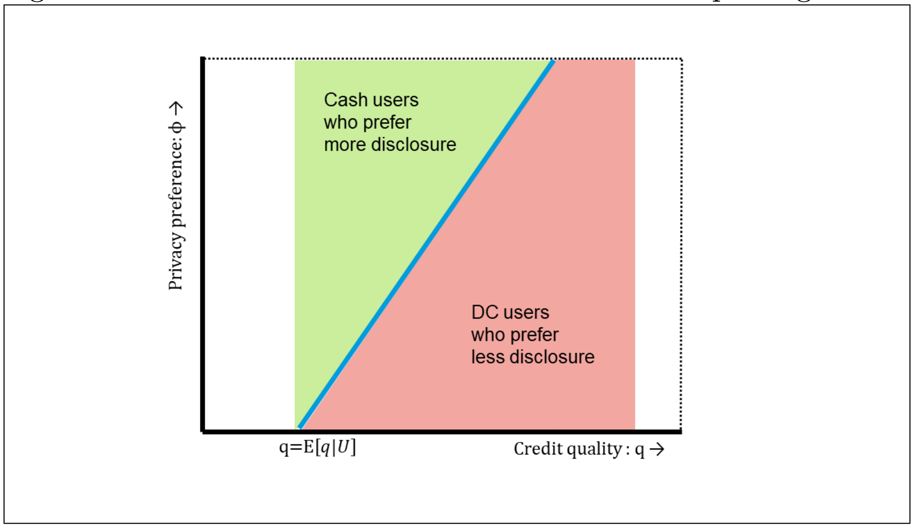
Additionally, equation (49) shows the probability of household revelation that emerges from the equilibrium found in Proposition 1 and which applies to the equilibrium of the quadratic extension (Appendix B.3) that is our point of comparison here. In the discussion below, we refer to the quadratic extension of the baseline model as the "comparator model".

$$\begin{array}{c}\text{Comparator model}\\ \text{equilibrium revelation probability}\end{array} = \begin{cases} 1 & \text{if } \frac{c}{\alpha\varphi}\left(\frac{q}{E[q|u]}-1\right) \geq 1 \\ 0 & \text{otherwise} \end{cases} \tag{49}$$

Figure 8 compares (48) and (49) and provides a rough inference on the differences between the free data porting and data monopoly models. This inference is imprecise because it does not account for the fact that $E\left[q|\, u\right]$ in (48) is not equal to $E\left[q|\, u\right]$ in (49). A full comparison would require a rederivation of $E\left[q|\, u\right]$ for the data porting model, in the manner of the proof of Proposition 1, and a subsequent numerical comparison of the differences between (48) and (49). This is beyond the scope of the current extension but the intuitions based on the rough inference are simple and, we conjecture, likely to carry over to a more complete analysis.

Figure 8 starts from the indifference frontier between cash and DC, which characterizes sorting in the comparator model. For households with $q < E\left[q|\, u\right]$, zero disclosure is privately optimal, because revelation comes with a privacy cost but cannot lead to lower customized loan rates. Such households choose $\theta_k = 0$ in the data porting model, matching the zero revelation that comes with cash in the comparator model. Instead, all households with $q > E\left[q|\, u\right]$ would like to choose an extent of disclosure, as can be seen from (48). When forced to choose between unrevealing cash and a fully revealing DC, the households that have $q > E\left[q|\, u\right]$ and are to the left of the indifference frontier opt for cash. Such households engage in partial revelation, $\theta_k \in (0, 1)$, in the data porting model, but are brought to

Figure 8: Disclosure in the baseline versus the data porting model.



zero disclosure in the comparator model. Instead, among the households to the right of the indifference frontier, there are many who, while they choose the DC when forced to choose between it and cash, would rather engage in less than full disclosure. Figure 8 purposely shows the "overdisclosure" zone as being larger than the "underdisclosure" zone. This is what the DC issuer plays on in the comparator model. Through cascades, the mass of DC users and, consequently, of revealed households increases when the DC becomes more intrusive, leading more households into overdisclosure, which the DC issuer commercializes.

As we are focusing on the good state, we know that $\theta^* = 0$: it would be socially optimal to ban the DC and any data porting. This means that the free data porting market is, from an aggregate welfare perspective, overdisclosing. But the private data monopoly, by actively playing on cascades and forcing households to a choice between full and no revelation, induces even more overdisclosure.

B.7 Intermediate y in the baseline model

This appendix analyzes the "Intermediate y" case of Table 1. From Sections 4.1 and 4.2, 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, per Proposition 8. Moreover, there can be a threshold $\widetilde{\theta}$, where $\theta > \widetilde{\theta}$ leads to an equilibrium as in Section 4.2 and $\theta < \widetilde{\theta}$ leads to an equilibrium as in Section 4.1.[137] In particular, at $\theta = 0$, the loan market for the unrevealed functions for any $y > \frac{4}{3}c$. Hence, $\widetilde{\theta} \in (0,1)$ exists if at $\theta = 1$ the unrevealed do not obtain credit: this can happen because $\frac{\partial E[q|u]}{\partial \theta} < 0$ from (26).

Proposition 8 *When $y \in \left(\frac{4}{3}c, 2c+\tau\right)$, then the equilibrium can be either as in Proposition 1 or as in Proposition 2, depending on parameters. For some parameterizations, there is 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. First, from Proposition 2, for $y = \frac{4}{3}c$ the loan market is shut to the unrevealed even for $\theta = 0$. Then, for $y = \frac{4}{3}c + \varepsilon$ the same must be true for some $\theta > 0$, because $\frac{\partial E[q|u]}{\partial \theta} < 0$.[138] This implies that there can be a threshold value of θ, $\widetilde{\theta}$, above which the

[137] Here, $\widetilde{\theta}$ can be found by solving $E\left[q|\, u\left(\widetilde{\theta}\right)\right] y = c$: this identifies the point where the home lender can expect zero profit from lending to the unrevealed market at the highest rate that borrowers are willing to accept, y. Moreover, the expression for $E\left[q|\, u\left(\widetilde{\theta}\right)\right]$ comes from replacing θ with $\widetilde{\theta}$ in (26). Numerical examples for $\widetilde{\theta}$, as well as associated welfare, have been calculated and are available on request.

[138] 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 the Proof of Proposition 2, the comparison becomes: the credit quality of the 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 $\tau > 0$.

outcome of Proposition 2 prevails and below which we get the outcome of Proposition 1.[139] In cases where there is a $\widetilde{\theta} \in (0,1)$, the outcome will be as in Proposition 2, since the DC issuer continues to optimally set $\theta = 1$.[140] ■

B.8 Expressions for the social planner with two instruments

This appendix records the main expressions for the two instrument case discussed in Section 6.2. We here show only the expressions themselves and not the underlying derivations (available on request). Households' payoff function can be written as

$$u(q, \varphi) = q \max\{(y - \tau I - R), 0\} - \alpha\varphi\Theta + JT + \text{transfers} \tag{50}$$

where J is an indicator variable that takes value 1 if the household chooses to use the DC and 0 if it chooses cash. $T > 0$ is the tax ($T < 0$ is the subsidy), the total revenues from which are remitted back (total costs of which are charged) to households lump-sum through the "transfers" term. E.g., a cash using household benefits from the presence of a tax, because DC users pay it, while the revenues are transferred back to all households. However, as transfers are independent of individual actions (given atomistic households), they do not affect household optimization, whereas the term JT does. The equilibrium sorting conditions for choosing DC over cash in Propositions 1 and 2, now become, respectively, $\varphi \leq \frac{1}{\alpha}\left(c\left(\frac{q}{E[q|u]} - 1\right) - \frac{T}{\theta}\right)$ and $q \geq \frac{1}{y-\tau}\left(\alpha\varphi + c + \frac{T}{\theta}\right)$. Furthermore, the expressions for

[139] 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 (17) 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 $\alpha\varphi < q(y - \tau) - c$ (from (31)). 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 $q\left(\frac{c}{E[q|u]} - \frac{c}{q} - \tau\right) < q(y - \tau) - c$ which becomes $\frac{c}{E[q|u]} < y$. With $E[q|u] \in \left[\frac{1}{2}, \frac{3}{4}\right]$ 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.

[140] This follows directly from the optimality of $\theta = 1$ in both Proposition 1 and Proposition 2.

$E\left[q|\,u\right]$ and $E\left[W\right]=\gamma W_h+\left(1-\gamma\right)W_l$ are

$$
\begin{aligned}
E\left[q|\,u\right] &= \frac{3c\left(2\theta-1\right)+\sqrt{3}\sqrt{\theta\left(\left(4\theta-3\right)\left(3T^2+6\alpha T\theta+4\alpha^2\theta^2\right)+3c^2\theta\left(1-\theta\right)\right)}}{\left(6c^2-8\alpha^2\right)\theta^2-6T\left(T+2\alpha\theta\right)}c \\
W_h &= \frac{3}{4}y_h-c-2\alpha\left(\theta-\frac{\theta\left(4\alpha+3c\right)+3T}{3c}E\left[q|\,u\right]\right) \\
W_l &= \begin{cases} \frac{(T-\theta(y_l-\tau-c))^2(T(2y_l-\tau)+\theta(y_l(y_l-c)+\tau(2c-\tau)))}{6\alpha\theta^2(y_l-\tau)^2} & \text{if } y_l<2\alpha+c+\tau+T \\ \frac{\theta^2\left((y_l-2\tau)\left(4\alpha^2+3(y_l-c)^2-6\alpha(y_l-c)\right)+3\tau^2(y_l-2\alpha-2c)\right)-3T^2y_l-6T\theta(\alpha+c)\tau}{3\theta(y_l-\tau)^2} & \text{if } y_l>2\alpha+c+\tau+T \end{cases}
\end{aligned}
$$