Black Box Credit Scoring and Data Sharing

Alessio Ozanne*

November 1, 2024

[Link to the Latest Version]

Job Market Paper

Abstract

Should credit scoring algorithms be transparent or opaque? I explore this question within a model where the lender uses data shared by borrowers for pricing and rationing credit, and is privately informed about how the algorithm maps data to allocations. I show that revealing the algorithm's parameters makes it vulnerable to gaming in the form of strategic withholding of unfavorable information. Under opacity, data withholding emerges as a hedging strategy against the unpredictability of the black box and the risk of credit rationing. The lender's optimal transparency regime maximizes data collection and is socially inefficient as it results in excessive credit rationing. Algorithmic opacity is often efficient as it reduces the stigma around data withholding, thereby expanding credit access for privacy-concerned borrowers. I analyze the distributional impacts of recent algorithmic transparency regulations and offer policy recommendations.

Keywords: FinTechs, Data, Disclosure, Algorithms, Opacity.
JEL Classification: D82, G14, G21, G38.

*Toulouse School of Economics. e-mail: alessio.ozanne@tse-fr.eu. I am grateful to Alexander Guembel, Matthieu Bouvard, Milo Bianchi, and Ansgar Walther for their guidance and support throughout the project. For helpful comments, I thank Basil Williams, Savitar Sundaresan, Christian Laux (discussant), Ulrich Hege, Sophie Moinas, Silvia Rossetto, Andrea Attar, Alex Smolin, Andrew Rhodes, Anna Sanktjohanser, Alexandre de Cornière, Christophe Pérignon, Gilles Chemla, Stefano Lovo, Simon Mayer, Stéphane Verani, Michele Bisceglia, Enrico Mattia Salonia, and participants in the 51st EFA Annual Meeting Doctoral Tutorial, EEA-ESEM 2024 Congress, 8th HEC Paris Finance PhD Workshop, 29th ENTER Jamboree. I thank Imperial College Businesses School, where part of this work was conducted, for the hospitality. This paper has been funded by the Agence Nationale de la Recherche under grant ANR FINRIS (ANR-20-CE26-0010).

1 Introduction

Motivation Recent advances in digital technology have transformed the way firms store, transmit, process, and analyze information. Notably, a growing number of financial institutions have embraced automated machine learning (ML) algorithms for the task of credit scoring and underwriting,[1] leveraging vast datasets that encompass both traditional financial metrics and alternative data sources.[2] ML algorithms, particularly when fed with alternative data, often reveal surprising relations between credit risk and seemingly unrelated variables, making it challenging to understand which specific factors influence algorithmic decisions, and how. This phenomenon is commonly known as the "black box" problem of artificial intelligence (AI).[3]

Algorithmic opacity has raised significant public concern, fostering uncertainty and mistrust in the lending process even with accusations of discriminatory practices,[4] and prompted individuals to conceal their online and offline behaviors, thereby restricting the information available to lenders.[5] On the other hand, making the algorithms' inner working transparent would potentially undermine their predictive ability, enabling borrowers to game the system through strategic information sharing.[6] Both regimes impact borrowers' willingness to share data — a crucial asset in a market plagued by asymmetric information — and may have a first-order impact on the allocative efficiency of credit markets. Should credit risk algorithms be transparent or opaque when we account for strategic information sharing? Regulators around the globe have decided to unpack the black box mandating transparency,[7] citing principles of truth, fairness, and equity, but possibly overlooking the importance of market efficiency.

[1] A prominent example is FinTech's "3-1-0 model": three minutes to apply, one second to approve, and zero human intervention. Traditional banks — such as J.P. Morgan and Bank of America — and credit scoring systems — such as the FICO and Vantage Score — have also started to employ automated algorithms. Automated credit underwriting is mainly applied to (unsecured) consumer and small business loans, as opposed to mortgages.

[2] Alternative sources of data include, but are not limited to, social media data, digital footprints, web searches, mobile phone usage patterns, e-commerce transactions, payment information, gig economy income, utility bill payments, childcare payments, and data from IoT devices. To put it in the industry's words, "all data is credit data" (Douglas Mirrell CEO of ZestCash to the New York Times). Alternative data have demonstrated superior performance compared to conventional FICO scores (see Iyer, Khwaja, Luttmer and Shue (2016), Bazarbash (2019), and Berg, Burg, Gombović and Puri (2020).

[3] See Pasquale (2015) for an investigation into our "Black Box Society".

[4] See Bartlett et al. (2022) for empirical evidence on discriminatory lending practices by FinTech lenders, and for a recent controversy see Neil Vigdor, "Apple Card Investigated After Gender Discrimination Complaints", The New York Times, November 11, 2019. See Glickson and Woolley (2020) for a study on human trust in AI.

[5] Sebastian Siemiatkowski, CEO of Klarna AB, once said: "Facebook is only as valuable as the information a consumer is willing to share and whether that consumer is willing to connect the financial service to their Facebook data". (See Evelyn M. Rusli, "Bad Credit? Start Tweeting", Wall Street Journal, April 1, 2013.))

[6] In a 2014 letter to the Consumer Financial Protection Bureau (CFPB), Freddie Mac raised concerns that disclosing their automated decision-making algorithm could make it vulnerable to being reverse-engineered (letter available here). See also Suzanne Woolley, "How More Americans Are Getting a Perfect Credit Score", Bloomberg, August 14, 2017.

[7] See the AI Act, Digital Services Act, Digital Markets Act of the European Commission; the Fair Credit Reporting Act and Equal Credit Opportunity Act of the Federal Trade Commission.

Overview This paper explores the lender optimal and the socially efficient level of transparency for credit risk algorithms, in light of the information that borrowers are willing to share with a financial institution. The model features a lender and a continuum of borrowers in need of cash to finance a project. Borrowers are on average creditworthy but differ in credit risk, with some having negative NPV projects. Additionally, borrowers possess data that can predict their probability of default. According to recent data regulations — such as Open Banking initiatives, the General Data Protection Regulation (GDPR), and the Payment Services Directives (PSD, PSD2) — this data remains private unless borrowers voluntarily share it with the lender. Furthermore, some borrowers prioritize privacy and consistently withhold information, regardless of what it might reveal about them.

The black box nature of the screening process arises from how the lender infers credit risk from data. The data-generating process (DGP) is unknown, while the lender can estimate its parameters from a previously collected training dataset, leveraging proprietary statistical technology. The allocation algorithm is calibrated on these privately known estimates, so the lender has exclusive knowledge of the mapping from disclosed data to allocations. The DGP is assumed to be characterized by a single parameter, representing the correlation between credit risk and an explanatory variable. Hence, evidence can be represented on a normalized interval of the real line, and while borrowers may not discern good from bad news, they can identify extreme evidence based on its distance to the mean. Within this framework, the paper investigates borrowers' voluntary data-sharing decisions under two scenarios: when the parameters governing the algorithm are kept secret (opacity) and when they are publicly known (transparency). Additionally, it explores the welfare and redistributive implications of algorithmic opacity and transparency, including the regime that maximizes the lender's profits, borrowers' surplus, and overall social welfare.

Preview of the Results The algorithm allocates credit in an intuitive way: high-risk borrowers are denied credit, while interest rates rise with credit risk. These rates reflect the competitive market rate, with an additional markup that grows with the lender's bargaining power. Still, the algorithm also converts evidence in credit-risk estimates and thus the map from data to allocations depends on the correlation between data and credit risk. If a borrower's characteristics positively correlate with credit quality, higher values of that variable lead to lower interest rates, while values below a certain threshold are denied credit. A symmetric reasoning applies when correlation is negative and higher values are progressively stronger signals of credit risk. The algorithm is influenced by the strength of the correlation, not just its sign. As the correlation decreases in absolute value, the algorithm rations credit to a narrower range of data realizations, and the interest rate schedule flattens.

When the algorithm's parameters are publicly known, borrowers may gain by withholding bad news while selectively sharing good news. This behavior, enabled by transparency, will be referred to as *gaming*. Gaming occurs when privacy-concerned borrowers are sufficiently numerous. In this case, high-risk borrowers can mimic their behavior and withhold data to obtain more favorable contractual terms. This happens because the stigma associated with data withholding is minimal as withholding is most likely non-strategic and therefore not a strong indicator of poor credit quality. As a result, credit-unworthy borrowers obtain credit by withholding information, while marginally creditworthy borrowers can secure lower interest rates. Conversely, borrowers share positive evidence because understanding the lender's statistical technology, they anticipate being classified as low risks. The number of privacy-concerned borrowers reflects the extent of strategic borrowers' gaming ability: as this number increases, more borrowers can secure better contractual terms by withholding information. Gaming does not occur when borrowers are primarily strategic and data sufficiently informative. In this case, data withholding becomes a strong indicator of credit risk and leads to credit denial. As a result, only (strategic) observationally creditworthy borrowers are funded and interest rates reflect their observable credit risk.

When the algorithm's parameters are kept secret, borrowers do not know the correlation between data and credit risk and data-withholding emerges as a prudent strategy against the algorithm's unpredictability. Although borrowers are risk-neutral, I term this strategic behavior as *hedging*, because it reflects their attempt to minimize exposure to uncertainty. Hedging is complete — meaning that all evidence is withheld — when the lender's bargaining power is sufficiently strong. In this scenario, the lender offers loans to all borrowers at a uniform rate, as the overall pool of borrowers is on average creditworthy. If data were shared, borrowers would expose themselves to the risk of credit rationing — with extreme borrowers being the most exposed — and conditionally on providing credit, the lender would capture most of their surplus through high personalized interest rates. Conversely, when borrowers have greater bargaining power, extreme evidence is disclosed in equilibrium, with the extent of disclosure increasing in the borrower's bargaining power. Although extreme borrowers are most at risk of credit rationing, they also stand to gain the most from favorable interest rates if they manage to secure credit. Borrowers with less conclusive evidence lack this upside potential and thus stick with the hedge of no disclosure. Overall, the borrowers' relative bargaining power modulates their risk-taking behavior when sharing data with the black box, shifting their decisions from withholding all information to disclosing the most conclusive evidence.

Interestingly, privacy-concerned borrowers receive markedly different treatment depending on whether the algorithm is transparent or opaque. Under transparency, they experience negative externalities due to the disclosure decisions of strategic borrowers. When strategic borrowers are numerous, privacy-concerned borrowers are excluded from credit due to the stigma surrounding data withholding, despite being, on average, creditworthy. This situation does not occur when the algorithm is opaque, as there is no stigma associated with data withholding. Since borrowers lack insight into what signals good or bad news, the lender cannot make adverse inferences based on a borrower's choice to withhold information. As a result, privacy-concerned borrowers, along with strategic non-disclosing borrowers, are able to secure credit.

Both transparency regimes lead to a data loss — either through strategic gaming or hedging behavior — which can undermine the algorithm's predictive accuracy and affect allocative efficiency. As a consequence, the transparency regime significantly influences welfare measures, including the lender's profits, borrowers' welfare, and overall social surplus. The lender's transparency choices maximize data extraction, focusing on effective credit rationing while forgoing the potential benefits of price personalization. This is due to the lender's risk neutrality: detailed credit quality information reduces variance but does not impact the ex-ante expected profits from creditworthy borrowers. Credit misallocations increase with borrowers' gaming ability (measured by the fraction of privacy-concerned users) and their hedging motives (captured by the lender's bargaining power). Therefore, the model suggests that an opaque algorithm should be the preferred choice for financial institutions in competitive credit markets with diffused privacy concerns, as hedging motives are limited while gaming more prevalent. Conversely, for lenders in concentrated credit markets with lower privacy concerns, transparency should be the optimal choice.

The lender's transparency choices are socially inefficient, resulting in over-collection of data and insufficient credit provision. Private and social trade-offs over accuracy and inclusion differ as the lender only partially internalizes the surplus generated by credit provision, resulting in excessive rationing from a social perspective. When the misalignment between the lender's and planner's preferences is moderate, data collection still enhances welfare by helping exclude strategic borrowers with negative surplus projects. However, the planner tends to favor opacity more than the lender does. This primarily stems from the adverse effects that the stigma of data withholding imposes on privacy-concerned but creditworthy borrowers under transparency. Opacity suppresses such a stigma enabling them credit and providing a social benefit that is only partially internalized by the lender. As a consequence, transparency is excessive. When the misalignment over credit provision increases, data collection starts to undermine welfare, as the borrowers that the lender seeks to ration hold positive-surplus projects. In this scenario, the welfare-optimal regime minimizes data extraction, making the lender's choices diametrically opposed to efficiency.

Outline The remainder of the paper proceeds as follows. Section 2 discusses the related literature. Section 3 describes the model. Section 4 derives the optimal allocation rule. Sections 5.1 and 5.2 present respectively the data sharing-equilibrium under transparent and opaque algorithms. Section 6 is dedicated to the welfare comparison of transparency regimes. Sections 7 and 8 discuss empirical and policy implications. All the proofs are relegated to the Appendix.

2 Related Literature

FinTech This paper connects to the literature on digital disruption in banking (see Vives (2019) for a review). Studies show that alternative data can outperform traditional metrics in predicting loan defaults (see Iyer et al. (2016), Bazarbash (2019), and Berg et al. (2020)), and expand credit access to individuals without a credit history (see Agarwal et al. (2019) and Gambacorta et al. (2020)). However, the ML methods that process these data can produce unintended distributional effects by capturing structural relationships among variables such as race, income, and gender (see Fuster et al. (2022)). I contribute to this literature by identifying a new channel through which these statistical technologies affect distribution and welfare: the transparency regime they adopt. While some studies have theoretically explored the optimal design of these technologies (see, e.g. Huang (2023), or He, Huang and Parlatore (2024), or Blickle et al. (2024)), I examine whether they should be transparent or opaque.

Voluntary Disclosure The paper is related to the large literature on voluntary information disclosure. The classical "unraveling argument" put forward by Grossman (1981) and Milgrom (1981) shows that, in equilibrium, a privately informed sender will disclose all the available information to avoid the receiver's averse equilibrium inference. The subsequent literature tried to explain the commonly observed lack of disclosure by introducing disclosure frictions.[8] Disclosure frictions in my model resemble those in Dye (1985) (uncertainty about evidence, or privacy cost in my model) and Bond and Zeng (2022) (uncertainty about audience, or statistical technology in my model). My paper contributes to the literature by comparing these frictions in the specific context of the credit market and conducting a comprehensive welfare analysis. In Bond and Zeng (2022), hedging behavior stems from the sender's risk aversion. In my model, the sender is risk-neutral, while risk-averse or risk-loving behaviors arise endogenously due to the specific market context considered. Unlike Bond and Zeng (2022), I also explore the welfare implications of hedging behavior.

Economics of Data and Privacy This work belongs to a growing literature on data, privacy, and their implications for markets (see Acquisti, Taylor, and Wagman (2016) and Bergemann and Bonatti (2019) for surveys). A portion of this literature explores the welfare impact of data regulations that grant customers rights over the sharing of personal data, such as the GDPR and Open Banking initiatives. For instance, Ali, Lewis and Vasserman (2023) study price discrimination under the GDPR, while He, Huang and Zhou (2023) study competition in the lending market under Open Banking, both yielding contrasting results

[8]These frictions include disclosure costs (Verrecchia (1983)), uncertainty about the evidence available to the sender (Dye (1985) and Jung and Kwon (1988)), information processing costs (Fishman and Hagerty (2003)), receiver's outside information (Harbaugh and To (2020) and Quigley and Walther (2024)) and uncertainty about audience's preferences (Bond and Zeng (2022)).

regarding consumer welfare. My work complements these studies by highlighting how data regulation interacts with algorithm transparency regulation, suggesting that both should be assessed jointly by regulators.

Gaming and Opacity The paper is related to the literature in machine learning, economics, and finance studying the manipulation of transparent scoring systems.[9] This literature typically considers types of data that can be distorted by agents at some cost, while I consider data as hard information that can be shared or withheld, but not misreported. A common finding in this literature is that an optimal scoring rule should underutilize data to deter manipulation. I show that opacity can emerge as an alternative strategy to data under-utilization.
A different strand of the literature has studied the role of opacity in softening strategic agents' gaming behavior in various contexts, including incentives schemes (Ederer, Holden and Meyer (2018)), stress testing (Leitner and Williams (2023)), and algorithmic decision-making (Wang et al. (2023) and Sun (2021)). The paper closest to mine is Sun (2021), which examines efficient algorithmic disclosure in a competitive lending market where borrowers can modify a binary feature at a cost. My paper is different in several aspects. First, I include interest rates and show that data-driven pricing exacerbates borrowers' gaming behavior. Second, I introduce the lender's bargaining power, leading to new hedging behavior in borrowers' data provision and enabling an analysis of the lender-optimal transparency regime. Third, I consider voluntary data-sharing in line with current Open Banking regulations. Finally, while Sun (2021) indicates that transparency harms the credit market, I highlight its potential benefits in certain contexts, aligning with the current policy debate.

Economics of Algorithms A nascent literature in economics studies how to regulate algorithmic decision-making when private and social objectives diverge (see Korinek and Balwit (2022) for a study of incentives misalignments in AI). A common scenario involves the algorithm aiming to maximize predictive accuracy, while the regulator seeks to balance accuracy with fairness,[10] typically by imposing restrictions on inputs or the prediction function (see, e.g., Kleinberg et al. (2018), Athey, Bryan and Gans, (2020), Rambachan et al. (2021), Blattner, Nelson and Spiess (2024), Liang et al. (2024)). My work contributes to this literature by considering a scenario in which the regulator prioritizes both accuracy and credit inclusion, placing a greater weight on the latter than the lender does. I consider the algorithm's transparency regime as being the only policy tool available to the regulator, in light of recent regulation on algorithmic transparency.

[9] See Hardt, Megiddo, Papadimitriou and Wootters (2016) for a classic paper in machine learning, Frankel and Kartik (2019), Frankel and Kartik (2021), Björkegren, Blumenstock and Knight (2021), Ball (2024), Perez-Richet and Skreta (2022), and Ekmekci et al. (2022) for recent papers in economics, and Cohn, Rajan and Strobl (2024) and Gamba and Hennessy (2024) for recent finance applications.

[10] See Cowgill and Tucker (2019) for an economic perspective on algorithmic fairness.

3 The Model

Economy The economy is composed of a lender and a continuum of penniless borrowers in need of cash to finance a project. Both players are risk-neutral and the borrowers are protected by limited liability. Each borrower has a project that requires a unit of cash and generates a return $X \in \mathbb{R}_+$ with probability $\theta \in [0,1]$ and generates nothing otherwise. The projects' probability of success is heterogeneous across borrowers and uniformly distributed, $\theta \sim U[0,1]$. I sometimes refer to θ as the borrower's credit quality, and to $1-\theta$ as the borrower's credit risk. If financed, the project generates a private benefit $b \in \mathbb{R}_+$ to the borrower, independently of whether the project succeeds or not. One can think of b as being cash flows that cannot be pledged to the lender, for example because of agency frictions (see Holmström and Tirole (1997)).

Assumption 1 (Positive NPV)**.** $\mathbb{E}(\theta)X - 1 > 0$.

Assumption 1 posits that the pool of borrowers is, on average, creditworthy (even excluding their private benefit, b). This implies that an uniformed lender finds it optimal to provide credit. However, since the return in the case of success, X, is finite, the lender aims to exclude from credit a subset of borrowers with high credit risk.

Data The borrower's quality θ is unknown to both the lender and the borrower at the time of contracting and can be predicted using data $z \in [0,1]$. The data-generating process (DGP) is an extension of the truth-or-noise information structure in Lewis and Sappington (1994), allowing for a negative correlation between the signal z and the state θ. Specifically, the DGP is such that

$$
\begin{aligned}
z &= \begin{cases} \theta & \text{with pr. } \lambda \\ \varepsilon & \text{with pr. } 1-\lambda \end{cases} \qquad \text{if } \lambda \geq 0, \\
z &= \begin{cases} 1-\theta & \text{with pr. } |\lambda| \\ \varepsilon & \text{with pr. } 1-|\lambda| \end{cases} \qquad \text{if } \lambda < 0,
\end{aligned} \tag{1}
$$

where $\varepsilon \sim U[0,1]$ is independent of θ and $\lambda \in [-1,1]$. Data perfectly reveals the borrower's credit quality with probability $|\lambda|$ and is noise, unrelated to quality, with the residual probability. Moreover, when $\lambda > 0$ (resp., $\lambda < 0$), data is positively (resp., negatively) correlated with credit quality. Hence, $|\lambda|$ will be referred to as the predictive power of data, while λ as the correlation between data and credit quality. The predicted credit quality conditional on data z is a weighted average of signal and noise with weights proportional to λ (see the proof of Lemma 1 for details):

$$\mu_\lambda(z) \triangleq \lambda\, z + (1-\lambda)\, \tfrac{1}{2} = \mathbb{E}(\theta|z). \tag{2}$$

Importantly, I assume that $\lambda \sim U[-1, 1]$ and the lender privately know its realization. In other words, the lender, having access to statistical technology, has superior knowledge of the DGP and thus privately knows how to convert data into credit risk estimates. This in turn implies that the lender is privately informed about the map from data to allocations (see below).

Data Sharing To capture current policies mandating consumer and borrower control over their data,[11] the data realization z is assumed to be private information of the borrower unless disclosed to the lender. Specifically, knowing z, the borrower chooses a message $m \in \{\varnothing, z\}$, where $m = z$ means sharing data with the lender, while $m = \varnothing$ means withholding data (as in He, Huang and Zhou (2023) and following the voluntary disclosure literature, e.g. Milgrom (1981), Grossman (1981)). The binary nature of the data sharing technology captures the fact that data is hard information: it can be transferred or withheld, but not modified or manipulated.[12] Building on Dye (1985) and Jung and Kwon (1988), a fraction $\pi \in (0, 1)$ of borrowers is assumed to withhold data for non-strategic reasons. These agents can be thought of as being privacy-concerned, lacking a credit history, being unbanked, technology-averse, or not possessing a device.[13] For concreteness, in what follows I assume them to be privacy-concerned. Importantly, the lender cannot distinguish privacy-concerned from strategic borrowers.[14]

Credit Allocation The credit allocation process is Nash bargaining and $\phi \in (0, 1)$ denotes the lender's bargaining power in the relationship. This parameter can be interpreted as reflecting the lender's market power,[15] arising from a more concentrated or segmented market structure.[16] Given the knowledge of the DGP, i.e. λ, and the borrower's disclosure decision

[11] Such policies include European Union's General Data Protection Regulation (GDPR) and Payment Service Directives (PSD, PSD2), California's Consumer Privacy Act (CCPA), China's Personal Information Protection Law (PIPL), and other globally adopted Open Banking Initiatives.

[12] This assumption differentiates this paper from previous studies (see Section 2, Gaming and Opacity), where data can be manipulated at some cost, and is more in line with Open Banking initiatives where information is usually stored in a relationship bank (Sharpe (1990) and Rajan (1992)), and can be transferred to a competitor at the customer's will, through an API, at no monetary cost.

[13] Studies show that agents' unwillingness to share information online depends on cultural factors (attitudes towards privacy, trust in institutions), demographic factors (age, education level, income level), comfort with technology, user experience (ease of sharing data), trust in the lender/firm (reputation, credibility, brand recognition), types of data shared (breadth, depth, sensitivity), etc. See, e.g. Acquisti, Taylor and Wagman (2016) for a general discussion and Morey, Forbath and Schoop (2015) for a survey.

[14] Lin (2022) empirically separate two components of consumers' privacy preferences: an intrinsic component (an exogenous taste for privacy, i.e. privacy concerns strictly speaking) and an instrumental component (an endogenous economic loss from revealing private information, i.e. strategic concerns). Goldfarb and Tucker (2012) show that intrinsic privacy concerns in digital markets have increased in recent years.

[15] More broadly, ϕ can also encompass other factors, independent of credit risk, that influence the lender's ability to charge higher interest rates, such as macroeconomic conditions (e.g., tight monetary policy) and regulations (e.g., looser capital requirements or consumer protection laws).

[16] As $\phi \to 1$ the market structure approaches a monopoly, while as $\phi \to 0$ it approaches perfect competition. This reduced-form approach to competition enables a focus on the effects of algorithmic disclosure on borrowers' data-sharing decisions, while abstracting from its potential strategic implications within firms. Moreover, modeling imperfect competition in the credit market with endogenous information acquisition is challenging

$m \in \{z, \varnothing\}$ the lender decides whether to provide creditor not and, conditional on providing credit, the lender also requires an interest rate to the borrower. Let $\ell \in \{0,1\}$ denote the credit rationing decision, with $\ell = 1$ denoting credit provision, and $x \in \mathbb{R}$ denote the gross interest payment.

Definition 1 (Algorithm)**.** *A credit allocation algorithm is* $a_\lambda(m) \triangleq \big(\ell_\lambda(m), x_\lambda(m)\big)_{m\in\{z,\varnothing\}}$.

I refer to $a \triangleq (\ell, x)$ as a credit allocation algorithm as, in equilibrium, it will be a map from data (or the lack of it) to allocations and it will be tailored to the specific DGP (described by λ). Hence, λ can be broadly thought of as parameters governing the lender's algorithm.

Algorithmic Transparency and Opacity In line with the Bayesian persuasion literature (see, e.g., Kamenica and Gentzkow (2011), Bergemann and Morris (2019), or Dworczak and Martini (2019)), I assume that the lender commits to a disclosure policy about the parameters that govern the allocation algorithm. Specifically, before learning λ the lender commits to reveal it or to conceal it.[17]

Definition 2 (Transparency and Opacity)**.** *The lender's algorithm* $a_\lambda(m)$ *is:*

- *transparent, if* λ *is public information,*
- *opaque, if* λ *is the lender's private information.*

Payoffs When the borrower's probability of success is θ and the lender employs an algorithm a, the lender's and the borrower's payoffs are, respectively:

$$\begin{aligned} V^B(\theta, a) &= \ell\left[\theta\big(X - x\big) + b\right], \\ V^L(\theta, a) &= \ell\left[\theta\, x - 1\right], \end{aligned} \tag{3}$$

while the social surplus (or the egalitarian social welfare) is:

$$W(\theta, a) = \ell\left[\theta X - 1 + b\right]. \tag{4}$$

and the literature often focus on simple algorithms with binary signal realizations (see, e.g., Broeker (1990)), with few exceptions (such as Blickle et al. (2024)). The reduced-form approach facilitates the study of more complex algorithms, aligning with advancements in ML technologies.

[17] The Bayesian persuasion literature typically studies optimal disclosure among the set of all the possible disclosure policies (i.e. state-contingent distributions over signal realizations). Here, I focus on a fully revealing and a fully concealing policy, in line with the ongoing debate regarding algorithmic transparency and opacity. It can be shown that, among symmetric monotone disclosure policies (i.e. symmetric convex partitions of the state space $[-1,1]$) it is without loss to restrict attention to policies that fully reveal extreme values $\lambda \in [-1, -\lambda^\circ) \cup (\lambda^\circ, 1]$ and pool intermediate ones $\lambda \in [-\lambda^\circ, \lambda^\circ]$ for some $0 \le \lambda^\circ \le 1$. I study the cases $\lambda^\circ \in \{0,1\}$, deferring a formal treatment of general disclosure policies to future work.

Timing and Equilibrium Concept The timing of the game is as follows:

$t = 0$ the lender commits to transparency or opacity, i.e. to reveal or conceal λ;

$t = 1$ (λ, θ, z) realizes, λ is observed by the lender (resp., everyone) under opacity (resp. transparency), z is observed by the borrower;

$t = 2$ the borrower shares or withholds data, i.e. he chooses $m(z) \in \{z, \varnothing\}$;

$t = 3$ the lender allocates credit, i.e. he chooses $\big(\ell_\lambda(m), x_\lambda(m)\big)_{m \in \{z, \varnothing\}}$;

$t = 4$ if the project is financed, the project's returns are realized.

The equilibrium concept is Perfect Bayesian Equilibrium.

Discussion of Model Assumptions A few comments about the model are in order.

- **Positive NPV** The lender extends credit based on prior beliefs, reflecting the common practice of incorporating alternative data sources in a secondary round of the screening process (see e.g., Nam (2024)), a practice referred to as *multistage screening*. Typically, borrowers who initially qualify for credit by submitting compulsory traditional metrics - such as credit scores and financial history - are then given the option to share additional information - such as payment data, social media activity, utility payments, or online behavior. This voluntary sharing can potentially enhance or impair their credit terms.

- **Unknown Credit Risk** Borrowers are often unaware of their own credit risk, a common assumption in the literature. In addition, this excludes the lender's ability to extract the borrower's private information through contractual terms, particularly collateral requirements.[18] In the realm of fintech credit, alternative data sources and advanced statistical technologies, typically serve as substitutes for traditional collateral.[19] Given that the focus of this paper is on the latter topic, I abstract from the possibility of designing truth-telling contracts.

- **Opaque Statistical Technology** The framework links algorithmic opacity to the underlying statistical technology. One can think of the DGP as being a priori unknown to everyone, while the lender, having access to a statistical technology and a training dataset, can obtain an estimate $\widehat{\lambda}$ of the DGP. Based on this estimate, the lender can tailor an algorithm, which can be described by the parameter $\widehat{\lambda}$. Private knowledge of the estimate of the DGP then coincides with private knowledge of the parameters governing

[18] An established body of literature has extensively examined the optimal design of screening contracts. See, e.g., Bester (1985), Freixas and Laffont (1990), and Besanko and Thakor (1987).

[19] Recent studies have explored the effectiveness of these alternative practices in addressing agency problems and expanding access to finance. See, e.g., Agarwal et al. (2019) and Gambacorta et al. (2020).

the algorithm. The model just assumes that such an estimate is unbiased, i.e. $\widehat{\lambda} = \lambda$. Essential to my model is the borrowers' uncertainty over of the estimate $\widehat{\lambda}$, while the underlying λ can be known or unknown to borrowers, and may even differ from $\widehat{\lambda}$. Such uncertainty reflects public mistrust in opaque ML technologies, originating, for example, from concerns about algorithmic fairness and bias.

- **Uniform Beliefs** Borrowers have uniform and symmetric beliefs over the employed statistical technology. This assumption captures the black-box nature of ML algorithms that utilize alternative data sources, as borrowers typically have little understanding of how these data points influence the lender's assessment of their credit quality. However, the assumption is mainly made to streamline results and can be relaxed, for instance assuming $\lambda \sim [\underline{\kappa}, \overline{\kappa}]$ where $-1 \leq \underline{\kappa} < 0 < \overline{\kappa} \leq 1$. Essential to borrowers' equilibrium hedging behavior against opacity is that they face some uncertainty about whether the algorithm treats data as good or bad news.

- **No Commitment to Allocation Rule** The lender does not commit to the employed algorithm. This is motivated by the common practice of *algorithm adaptability*, where systems are enabled to learn continuously and evolve based on ever-changing data inputs and economic conditions. Moreover, from a modeling perspective, commitment to a secret allocation rule seems contradictory, besides being sub-optimal. Once data is shared, the lender would optimally update the allocation rule, as deviations are not detectable, essentially leading to the above timing structure.

- **Commitment to Disclosure Policy** Committing to opacity is straightforward: the lender simply keeps the algorithm secret, regardless of its complexity. Commitment to transparency is more nuanced and may depend on the algorithm's *interpretability*. When using interpretable algorithms, such as logistic regression, the lender can directly disclose model parameters — akin to the λ in my model. However, when employing more complex ML techniques - such as support vector machines, tree-based models, gradient boosting machines, or deep learning methods - the lender can be, at least, *explainable*. In practice, this might include disclosing feature importance metrics, SHAP values, LIME explanations, partial dependence plots, and permutation importance to provide borrowers insights into how different features influence predictions (see, e.g. Molnar (2023)).

- **Single-Dimensional Data** The model considers a single explanatory variable, simplifying both interpretation and algebra, facilitating welfare analysis. This framework can be extended to include multidimensional characteristics with an all-or-nothing disclosure technology. In this setup, the probability of success θ is a weighted average of n latent characteristics $t = (t_1, \dots, t_n)$, i.e. $\theta = \sum_{i=1}^{n} \alpha_i t_i$ where $t_i \sim U[0,1]$ and $\alpha_i \in [0,1]$

with $\sum_{i=1}^{n} \alpha_i = 1$. Each latent characteristic t_i can be estimated from an observable z_i generated by a DGP analogous to Equation (1), resulting in predicted credit quality being equal to the hyperplane $\mu_\lambda(z) = \sum_{i=1}^{n} \alpha_i \mu_{\lambda_i}(z_i)$, where $\lambda = (\lambda_1, \ldots, \lambda_n)$ and $z = (z_1, \ldots, z_n)$. The disclosure technology allows borrowers to share all observables or none, i.e. $m(z) \in \{z, \varnothing\}$. Intuitively, under transparency, borrowers can identify which direction of the hyperplane constitutes bad news, allowing them to selectively withhold such information (gaming). Conversely, under opacity, borrowers know that evidence near the boundaries of the z space constitutes an extreme outcome, enabling them to withhold this evidence (hedging).

4 Data-Driven Credit Underwriting

This section describes the optimal credit allocation algorithm in the situation where the lender has access to the borrower's data z. This is also the allocation the lender chooses when the borrower chooses to share data voluntarily.

Given the lender's estimate of the borrower's credit quality $\mu_\lambda(z)$, the optimal interest rate x is chosen to maximize the Nash product:

$$\max_{x \in \mathbb{R}} \quad \Big[\mu_\lambda(z)\, x - 1\Big]^{\phi} \Big[\mu_\lambda(z)\big(X - x\big)\Big]^{1-\phi}, \tag{5}$$

and has to satisfy the lender's participation constraint $\mu_\lambda(z)x - 1 > 0$.[20] When the constraint cannot be satisfied, I say that the borrower is credit rationed,[21] i.e. $\ell = 0$. The Nash product in Equation (5) is a weighted geometric average of the lender's and borrower's surplus for a given interest rate x with weights proportional to their bargaining power, respectively ϕ and $1 - \phi$. Note that the borrower's private benefit b does not enter the Nash product, as these are cash flows that cannot be pledged to the lender. The optimal allocation rule is described in the following lemma.

Lemma 1 (Optimal Credit Allocation). *If the lender observes the borrower's data z, the optimal credit allocation algorithm $a_\lambda(z)$ is:*

$$\begin{aligned} x_\lambda(z) &= \frac{1}{\mu_\lambda(z)} + \phi\left(X - \frac{1}{\mu_\lambda(z)}\right), \\ \ell_\lambda(z) &= \mathbb{1}\left\{\mu_\lambda(z) > \frac{1}{X}\right\}. \end{aligned} \tag{6}$$

[20] This axiomatic (or cooperative) solution has a strategic microfoundation, as it coincides with the sequential equilibrium of a (non-cooperative) bargaining game with alternating offers (see Rubinstein (1982) and Binmore (1987)).

[21] I implicitly assume that when the lender makes nil profits she does not provide credit. This allows me to conveniently write the no disclosure set (analyzed in Section 5) as a closed set, thus improving readability. The assumption is otherwise inconsequential.

Proof. See the Appendix. ∎

The allocation rule is rather intuitive and depends on predicted credit quality as follows. First, the interest rate is decreasing in expected quality $\mu_\lambda(z)$: less risky borrowers get lower interest rates. Moreover, the interest rate equals the competitive rate $1/\mu_\lambda(z)$ plus a mark-up that is increasing in the lender's bargaining power ϕ, the return of the project X and expected quality $\mu_\lambda(z)$. Second, the credit rationing decision is a cutoff rule: borrowers with credit quality below the threshold $1/X$ are credit rationed.

Importantly, the algorithm depends on the correlation λ between data and credit quality, through the lender's forecast of quality $\mu_\lambda(z)$. When λ is positive (negative), higher data realizations are good news (bad news) and the interest rate decreases (increases) in z, while it is flat and independent of data when $\lambda = 0$. Moreover, the interest rate is steeper the higher the predictive power of data $|\lambda|$. Similarly, the allocation rule rations credit to different sets of data points depending on the value of λ. One can rewrite the credit rationing rule in Equation (6) as

$$\ell_\lambda(z) = \begin{cases} \mathbb{1}\big\{z > r(\lambda)\big\} & \text{if} \quad \lambda \in \big[\overline{\lambda}, 1\big] \\ 1 & \text{if} \quad \lambda \in \big(\underline{\lambda}, \overline{\lambda}\big) \\ \mathbb{1}\big\{z < r(\lambda)\big\} & \text{if} \quad \lambda \in \big[-1, \underline{\lambda}\big], \end{cases} \tag{7}$$

where

$$r(\lambda) = \frac{1}{2} - \frac{1}{\lambda}\left(\frac{1}{2} - \frac{1}{X}\right), \tag{8}$$

will be refereed to as the rationing threshold, while $\overline{\lambda} \triangleq 1 - 2/X \in (0,1)$ and $\underline{\lambda} = -\overline{\lambda}$. Note that when the correlation is sufficiently high, $\lambda \in \big[\overline{\lambda}, 1\big]$, the rationing threshold is positive, $r(\lambda) \in [0, 1/X]$, and some borrowers with low data points are credit rationed. Conversely, borrowers with high data points do not get credit when the correlation is sufficiently negative, $\lambda \in \big[-1, \underline{\lambda}\big]$ so that $r(\lambda) \in [1 - 1/X, 1]$. Finally, when the predictive power of data is low $|\lambda| < \overline{\lambda}$, there is no data-driven credit ratio; still, data will be optimally used to price credit (as long as $|\lambda| > 0$).

Given the optimal allocation rule, the lender's and the borrowers' expected profits given a data realization z are, respectively:

$$\begin{aligned} V_\lambda^L(z) &= \ell_\lambda(z)\Big[\phi\Big(\mu_\lambda(z)X - 1\Big)\Big], \\ V_\lambda^B(z) &= \ell_\lambda(z)\Big[(1-\phi)\Big(\mu_\lambda(z)X - 1\Big) + b\Big]. \end{aligned} \tag{9}$$

Conditional on credit provision, the lender and the borrower obtain a fraction of the (contractible) social surplus equal to their bargaining power, while the borrower also enjoys the private benefit. Moreover, greater statistical accuracy (higher $|\lambda|$) has a redistributive impact, creating winners and losers (as in Fuster et al. (2022)). Better-than-average borrowers enjoy lower interest rates and are thus better off as a result of greater predictive power. In contrast, riskier borrowers are worse-off, facing higher interest rates and potential credit rationing as a result of increased precision.

5 Voluntary Data Sharing

This section studies borrowers' voluntary data-sharing behavior. In this framework, the lender draws inferences from both the data provided and the borrower's disclosure decisions. The analysis considers in turn transparent and opaque allocation algorithms, as defined in Definition 2.

Before studying the borrowers' data-sharing decision, consider the lender's equilibrium inference following data withholding. Let $Q \triangleq \{z \in [0,1] : m(z) = \varnothing\}$ be the set of borrowers that strategically withhold data, the lender's inference of the borrower's credit quality is $\mu_\lambda(\varnothing) = \mu_\lambda(z(\pi, Q))$ where $\mu_\lambda(z)$ is defined in Equation (2), while

$$\begin{aligned} z(\pi, Q) &\triangleq \omega(\pi, Q)\tfrac{1}{2} + \big(1 - \omega(\pi, Q)\big)\mathbb{E}(\theta | \theta \in Q), \\ \omega(\pi, Q) &\triangleq \frac{\pi}{\pi + (1-\pi)\Pr(z \in Q)}. \end{aligned} \tag{10}$$

Data withholding can either be strategic, and stem from a borrower with data realization $z \in Q$, or be non-strategic. Thus, following withholding, the lender assigns an inferred data realization $z(\pi, Q)$ that is a weighted average of the expected data realization withheld by strategic and non-strategic borrowers. The weight $\omega(\pi, Q)$ is the posterior probability of non-strategic withholding. It increases with π, the fraction of non-strategic users, and decreases with the mass of strategically withheld data realizations. Note that, the inferred data realization $z(\pi, Q)$ is also weighted by λ, reflecting the likelihood of data being informative, rather than noise. Following withholding, the algorithm maps the inferred data quality to allocations as in Lemma 1 replacing $z(\pi, Q)$ to z (see Lemma A.1 in the Appendix for a formal treatment).

5.1 Gaming under Transparency

This section shows that transparency makes the algorithm vulnerable to gaming in the form of strategic withholding of unfavorable information. When the parameters governing the algorithm are publicly known, the borrower's data-sharing strategy depends on this knowledge and is formalized in the following proposition.

Proposition 1 (Data-Sharing to a Transparent Algorithm)**.** *When the lender uses a trans-*

parent algorithm, the set of strategic borrowers that optimally withhold data is:

$$\mathcal{G}(\lambda,\pi)=\begin{cases}\Big[0,\max\big\{r(\lambda);\gamma(\pi)\big\}\Big] & \textit{if } \lambda\in(0,1],\\ \Big[0,1\Big] & \textit{if } \lambda=0,\\ \Big[\min\big\{r(\lambda);1-\gamma(\pi)\big\},1\Big] & \textit{if } \lambda\in[-1,0),\end{cases}\tag{11}$$

where

$$\gamma(\pi)\triangleq\frac{\sqrt{\pi}(1-\sqrt{\pi})}{1-\pi}\in\left(0,\frac{1}{2}\right)\tag{12}$$

is increasing.

Proof. See the Appendix. ■

Consider the case where the correlation is positive and low, such that the algorithm uses data for pricing but not to ration credit, i.e. $\lambda\in(0,\overline{\lambda})$ (see Equation (7) and the discussion that follows). The algorithm is known to assign lower interest rates to borrowers with higher z. Risky borrowers — with low z — would get pretty high interest rates by sharing information, and thus opt to withhold data as this grants them better contractual terms. Suppose that the set of borrowers that withhold data is $Q=[0,q]$ for some $q\in(0,1)$. By pooling with non-strategic borrowers, who have an average credit quality of $1/2$, borrowers with $z\in Q$ are perceived as lower credit risks than they actually are and thus receive lower interest rates. However, data withholding is perceived by the lender as a signal of credit risk and better borrowers prefer to separate from the risky no-disclosure pool by sharing information.

The threshold q can be identified as the borrower that receives the same interest rates by sharing and withholding data. For this to occur, the inferred data realization following data withholding must match the borrower's actual data realization. The indifference condition can be expressed as

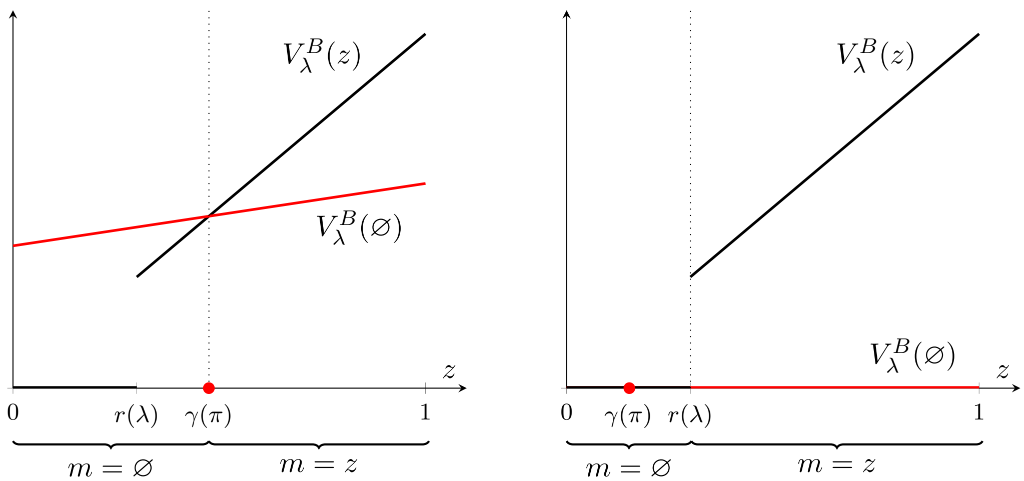
$$q=z(\pi,q),$$

where $z(\pi,q)$ is determined by substituting $Q=[0,q]$ into both lines of Equation (10). Its unique (positive) solution is $q=\gamma(\pi)$ given by Equation (12). Note that this threshold is independent of the predictive power of data λ. This is because the lender's equilibrium inference, $\mu_\lambda(\cdot)$, scales by the same weight λ both the disclosed data realization z and the data realization $z(\pi,q)$ inferred from data withholding. Hence, λ is irrelevant to the borrower's indifference condition, all that matters is the comparison between the borrower's actual data realization and the inferred data realization when data is withheld. On the other hand, the indifference condition depends on the fraction of non-strategic borrowers π. The indifference threshold $\gamma(\pi)$ is increasing and strictly between 0 and $1/2$. When borrowers are primarily strategic ($\pi\to 0$), data withholding is a strong indicator of low credit quality. Consequently, the algorithm significantly penalizes withholding by assigning very high interest rates. As a result, even the riskiest borrowers choose to separate, causing the no-disclosure set to shrink, and $\gamma(\pi)$ approaches zero. This essentially aligns with the unraveling result of by Grossman (1981) and Milgrom (1981), which states that all available information is disclosed to prevent an adverse equilibrium inference by the receiver. When the fraction of privacy-concerned borrowers increases, unraveling forces are milder because withholding data becomes a less reliable indicator of low credit quality. Consequently, a growing number of risky borrowers prefer to withhold information, as this strategy yields increasingly favorable interest rates as π rises. In the opposite limit, where borrowers are primarily non-strategic ($\pi \to 1$), withholding data provides no information, and the algorithm's equilibrium inference aligns with the prior. As a result, only above-average borrowers choose to separate, and $\gamma(\pi)$ approaches 1/2.

Consider now more data-intensive allocation rules such that $\lambda \geq \bar{\lambda}$. Disclosed data is used not only for pricing but also for rationing credit. In particular, borrowers with $z \in [0, r(\lambda)]$, with $r(\lambda) > 0$ given by Equation (8), do not receive credit if they share their data. However, withholding information may allow them to obtain credit if the inferred data realization $z(\pi, \gamma(\pi)) = \gamma(\pi)$ exceeds the rationing threshold $r(\lambda)$. This happens when unraveling forces are sufficiently mild that the lender's equilibrium inference is not too averse. Specifically, this occurs when non-strategic borrowers are relatively numerous, i.e. $\pi > \gamma^{-1}(r(\lambda)) = \left(r(\lambda)/(1-r(\lambda))\right)^2 \in (0,1)$ or, equivalently, when the predictive power of data is sufficiently low, i.e. $\lambda < r^{-1}(\gamma(\pi)) = \left(\frac{1}{2} - \frac{1}{X}\right) / \left(\frac{1}{2} - \gamma(\pi)\right) \in (\bar{\lambda}, 1)$. The equilibrium is depicted in Figure 1, Panel (a). When these conditions are not met, unraveling forces are stronger and the lender's equilibrium inference is so averse that withholding data induces credit rationing. As a consequence, borrowers with data below the rationing threshold $r(\lambda)$ are indifferent between sharing and withholding as they do not receive credit anyway. Only observationally credit-worthy borrowers share information and get funded at an interest rate that reflects their true expected credit risk. The equilibrium is depicted in Figure 1, Panel (b).

When $\lambda < 0$, symmetric reasoning applies and the right tail of the data distribution is withheld in equilibrium. When $\lambda = 0$, data is uncorrelated with quality and the algorithm extends credit ignoring any disclosed information. As a consequence, borrowers are indifferent between sharing and withholding data.

The equilibrium withholding of information can be thought of as a form of gaming — enabled by transparent allocation rules — by which strategic borrowers secure better contractual terms by selectively concealing bad news while sharing good news. The misallocation of credit resulting from gaming has welfare consequences for both the lender and the borrowers and will be analyzed in Section 6. Gaming can result in lower interest rates for risky but credit-worthy borrowers and even induce credit provision to credit-unworthy borrowers. In both cases, the extent of gaming increases with the fraction of non-strategic users, which thus can be seen as a measure of gaming ability. Moreover, gaming is more pronounced when data is more informative and the allocation rule relies more strongly on it.[22] This result is related to several papers that examine the manipulation of allocation rules by strategic users, finding that gaming incentives increase when the rule heavily relies on data (see the paragraph on Gaming and Opacity in Section 2). These papers also suggest that an optimal scoring rule should underutilize data to mitigate gaming. Section 5.2, explores an alternative strategy that lenders might employ to deter gaming: making the rule opaque without reducing its data sensitivity.



(a) $r(\lambda) < \gamma(\pi)$.

(b) $\gamma(\pi) \leq r(\lambda)$

Figure 1: **Gaming under Transparency.** The figure depicts the borrower's surplus from data sharing, $V^B_\lambda(z)$ (in black), and from data withholding, $V^B_\lambda(\varnothing)$ (in red), when $\lambda > \overline{\lambda} > 0$. The red dot corresponds to the borrower's data realization inferred from data withholding, and equals $\gamma(\pi)$ defined in Equation (12). $r(\lambda)$ denotes the rationing threshold and is defined in Equation (8). Panel (a) shows the case where privacy-concerned borrowers are sufficiently numerous, i.e. $\pi > \gamma^{-1}(r(\lambda))$, and gaming occurs in equilibrium. Panel (b) shows the case where privacy-concerned borrowers are few, i.e. $\pi \leq \gamma^{-1}(r(\lambda))$, and gaming does not occur in equilibrium.

5.2 Hedging under Opacity

This section examines borrowers' data-sharing behavior when the algorithm is opaque, meaning the parameters governing it are kept secret from borrowers. While opacity can sometimes

[22] When $\lambda \in [\overline{\lambda}, \widehat{\lambda}(\gamma(\pi)))$, the mass of credit-unworthy borrowers that get credit by withholding information increases with the predictive power of data.

mitigate gaming, it also induces borrowers to hedge against the unpredictability of the algorithm by withholding information. The equilibrium data-sharing strategy is described in the following proposition.

Proposition 2 (Data-Sharing to an Opaque Algorithm)**.** *When the lender uses an opaque algorithm, the set of strategic borrowers that optimally withhold data is:*

$$Q = \mathcal{H}(b,\phi) \triangleq [\eta(b,\phi), 1-\eta(b,\phi)]$$

where

$$\eta(b,\phi) \triangleq \max\left\{0, \frac{1}{X}\left(1-\frac{2b}{1-\phi}\right)\right\} \in \left[0, \frac{1}{X}\right) \tag{13}$$

is weakly decreasing in ϕ and b.

Proof. See the Appendix. ■

Before providing the intuition behind this result, it is helpful to understand why data-withholding constitutes a safe hedging strategy against algorithmic opacity. Consider *any* symmetric data-withholding set Q.[23] For any of these sets, the lender's inference of credit quality following data-withholding is

$$\mu_\lambda(\varnothing) = \mu_\lambda\left(\tfrac{1}{2}\right) = \tfrac{1}{2},$$

and is thus independent of i) the lender's statistical technology, and ii) the actual data-withholding set (see Equation (10)). No matter the true link between data and credit quality (i.e. λ) and the size of the no-disclosure pool, the data-withholding set contains borrowers with mixed positive and negative evidence, so the lender does not infer much from data withholding. As a consequence, as long as the no-disclosure pool is symmetric, there is no stigma attached to data withholding, and disclosing borrowers impose no externalities on withholding borrowers. Moreover, since the overall pool of borrowers is credit-worthy, the lender provides credit following no-disclosure. For similar reasons, the ex-post interest rate a borrower gets from withholding is independent of λ. It follows that, if the data-withholding set is symmetric, data-withholding is a safe strategy against algorithmic opacity: it insures the borrower against the risk of credit rationing and the variability in interest rates.

Consider now what the borrowers expect from data sharing. We can distinguish two types of borrowers depending on their data realizations:

- Borrowers with extreme data realizations $z \in [0, 1/X] \cup [1-1/X, 1]$ face a positive probability of being denied credit when they share information, and this probability increases

[23] A set $Q \subset [0,1]$ is symmetric about $\frac{1}{2}$ if $\forall z \in Q$, $1-z \in Q$. Some examples are: the entire interval $Q = [0,1]$, a subinterval of the form $Q = [q, 1-q]$ with $q \in [0, 1/2)$, disjoint subintervals of the form $Q = [0,q] \cup [1-q, 1]$ with $q \in (0, 1/2]$, etc.

as their data realization becomes more extreme — i.e. when z moves further from the center of the distribution. This is intuitive, as increasingly extreme data realizations are deemed credit-unworthy for progressively weaker correlation values λ, resulting in a denial of credit by a wider range of allocation rules that the lender may employ.[24] Conditional on getting credit, the interest rates are also uncertain — as these depend on λ through the lender's opaque statistical inference from data — and the expected interest rate decreases as data becomes more extreme. This is also intuitive, as, conditional on receiving credit, increasingly extreme data realizations are seen as progressively stronger signals of credit quality.[25]

- Borrowers with central data realizations $z \in (1/X, 1-1/X)$ face no risk of credit rationing and they always obtain credit by sharing information. Their data realizations are close to average and represent a weak indicator of credit risk regardless of whether the data is positively or negatively correlated with credit quality. Since the average borrower is creditworthy, they are granted credit by every allocation rule the lender might employ. However, these borrowers still face algorithmic opacity because the interest rate they receive is uncertain. Given the borrowers' risk neutrality and the symmetric treatment of data across rules, this uncertainty perfectly balances out, and the expected interest rate is independent of z.

The equilibrium data-sharing behavior can now be better understood. When the lender's bargaining power is relatively high (i.e. $\phi \geq 1 - 2b$) or, equivalently, when the borrower's private benefit is substantial (i.e. $b \geq (1-\phi)/2$), the equilibrium involves no data-sharing, i.e. $Q = [0, 1]$. In this scenario, borrowers prioritize securing credit and enjoying their private benefits b over obtaining marginally lower interest rates, as interest rates extract most of their surplus once the credit is extended. Consequently, the utility of data sharing exhibits a symmetric inverted U-shape: borrowers with more extreme data realizations face an increasing risk of credit rationing and would like to appear closer to the average than they truly are to mitigate this risk. This can be achieved by withholding information. The strategy labels them

[24] Consider $z \in [0, 1/X]$. From Equation (7) and Equation (8), a data realization in this set is denied credit when $z \leq r(\lambda)$, or equivalently when the correlation between data and credit quality is sufficiently high, i.e. when

$$\lambda \geq r^{-1}(z) \triangleq \widehat{\lambda}(z) = \frac{\frac{1}{2} - \frac{1}{X}}{\frac{1}{2} - z} \in (\overline{\lambda}, 1],$$

where $r^{-1}(\cdot)$ denotes the inverse function of $r(\cdot)$. The probability of getting credit is thus $p(z) \triangleq \Pr(\lambda < \widehat{\lambda}(z)) = (\widehat{\lambda}(z) + 1)/2$, so that the probability of credit rationing, $1 - p(z)$, is decreasing in z for $z \in [0, 1/X]$. Symmetric reasoning applies for $z \in [1 - 1/X, 1]$.

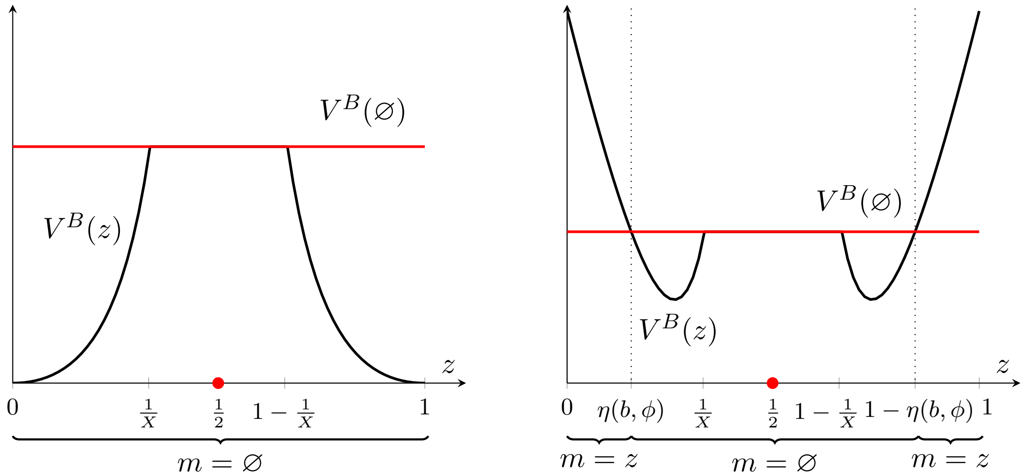
[25] Consider $z \in [0, 1/X]$. The inferred credit quality across all the allocation rules that provide the borrower credit is $\mu_-(z) = \mathbb{E}(\mu_\lambda(z) | \lambda < \widehat{\lambda}(z))$ and is decreasing in z for $z \in [0, 1/X]$. This is because the borrow receives credit when $\lambda < \widehat{\lambda}(z)$, i.e. when data is predominately negatively correlated with credit quality, so when lower data realizations are viewed as more favorable. Therefore, conditional on receiving credit, the expected interest rate increases with z. Symmetric reasoning applies for $z \in [1 - 1/X, 1]$.

as average borrowers, fully insulates them against the risk of credit rationing, and is therefore strictly optimal. In contrast, borrowers with central data realizations are indifferent between the two strategies, as they obtain credit anyway and are risk-neutral conditional on receiving credit. No borrower has an incentive to separate from the no-disclosure pool. Hence, data-withholding emerges as a safe hedging strategy against the unpredictability of the algorithm: it insures the borrower against the risk of credit rationing and the variability in interest rates.[26] The equilibrium is depicted in Figure 2, Panel (a).

Conversely, when the borrower has relatively more bargaining power (i.e. $\phi < 1 - 2b$) or a lower private benefit ($b < (1-\phi)/2$), the most extreme evidence is disclosed in equilibrium, i.e. $Q = [\eta(b,\phi), 1-\eta(b,\phi)]$ with $\eta(b,\phi) \in (0, 1/X)$. In this case, the potential for better interest rates drives extreme borrowers' decision to share information. These are still the most likely to face credit rationing if they disclose but also stand to gain the best rates if they manage to obtain credit. Consequently, when the potential upside reward outweighs the guaranteed benefit of obtaining credit, these borrowers forgo the safety of non-disclosure and take the risk of sharing information. However, less extreme borrowers have less exposure to the potential upside and therefore stick to the safe hedge. We can thus write the no-disclosure set as $Q = [q, 1-q]$ and identify the threshold q as the borrower for whom these forces perfectly balance out, being indifferent between sharing and withholding data. The indifferent borrower is given by $q = \frac{1}{X}\left(1 - \frac{2b}{1-\phi}\right) \in (0, 1/X)$. This threshold increases when both ϕ and b decrease: as borrowers gain bargaining power or derive lower benefits from securing credit, hedging motives decrease and they share more information in equilibrium. In the limit when b approaches 0 hedging motives are absent: no matter the lender's bargaining power, every borrower facing the threat of credit rationing strictly prefers to disclose information, hoping for a better interest rate. The equilibrium is depicted in Figure 2, Panel (b).

In summary, algorithmic opacity prompts some borrowers to hedge against its unpredictability by withholding information. Hedging motives are strengthened when the lender can more easily extract borrowers' surplus, and can even prevent data sharing from occurring in equilibrium. However, leveraging borrowers' risk-taking behavior, opacity may also lead to the extraction of extreme and highly conclusive evidence when borrowers possess greater bargaining power. Thus, algorithmic opacity presents both benefits and costs in terms of data collection and the resulting reduction of misallocation of credit. These factors will be studied in Section 6.

[26] Since the value from data-sharing is symmetric around $z = 1/2$, the set of borrowers withholding data, Q, is also symmetric in equilibrium. Note that the hedge is perfect because of the symmetry of Q, but is only partial when the conjectured Q is asymmetric. Still, the intuition extends to asymmetric no-disclosure sets and perfect hedging will be optimal in equilibrium. Similarly, a symmetric equilibrium data-withholding set is due to the symmetry in algorithmic uncertainty, which drives the symmetry of the value from data-sharing around $z = 1/2$. One can show that hedging type of behavior (i.e. pooling at the extremes) emerges in equilibrium when $\lambda \sim [\underline{\kappa}, \overline{\kappa}]$ where $-1 \leq \underline{\kappa} < 0 < \overline{\kappa} \leq 1$ and $\underline{\kappa}$ is sufficiently low.



(a) $\phi \geq 1 - 2b$.

(b) $\phi < 1 - 2b$

Figure 2: **Hedging under Opacity.** The figure depicts the borrower's expected surplus from data sharing, $V^B(z)$ (in black), and from data withholding, $V^B(\varnothing)$ (in red). The red dot corresponds to the borrower's data realization inferred from data withholding, and equals $\frac{1}{2}$. $\eta(b, \phi)$ denotes the indifferent borrower and is defined in Equation (13). Panel (a) shows the case where the lender's bargaining power is sufficiently high, and hedging is complete in equilibrium. Panel (b) shows the case where the lender's bargaining power is sufficiently low, and hedging is only partial in equilibrium.

5.3 Externalities and Privacy Concerns

Interestingly, privacy-concerned borrowers are affected differently depending on whether the lender's algorithm is transparent or opaque.

Corollary 1 (Externalities). *When the lender uses a transparent algorithm, non-strategic borrowers are denied credit when they are few, i.e. when* $\pi \leq \gamma^{-1}(r(\lambda)) \in (0, 1)$. *Instead, when the lender uses an opaque algorithm, non-strategic borrowers always obtain credit.*

Under transparency, when privacy-concerned borrowers are few, no-disclosure is a strong signal of low credit quality and leads to credit denial. Hence, strategic borrowers impose a negative externality on privacy-concerned borrowers through their data-sharing decisions, ultimately pushing them out of the market, despite being creditworthy on average. This operates through the classic unraveling logic and stems from the lender's inability to differentiate between strategic and non-strategic borrowers. On the contrary, unraveling forces are absent under opacity. The no-disclosure pool contains both observationally good and bad borrowers, thus there is no stigma attached to no-disclosure, no matter the true statistical link between data and credit quality. As a result, borrowers who disclose information impose no externalities on privacy-concerned borrowers, allowing them to always receive credit.

6 Welfare Analysis

The previous discussion highlighted that both transparency and opacity lead to information losses through either gaming or hedging behavior of strategic borrowers. This section studies the resulting misallocation of credit and mispricings. It demonstrates that transparency regimes have significant welfare implications, affecting the lender's profits, borrowers' welfare, and overall social surplus.

6.1 Lender's Profits

We begin by analyzing the transparency regime that maximizes the lender's ex-ante profits. Since data is valuable for the lender, the optimal regime minimizes the information loss resulting from borrowers' strategic behavior. The lender's transparency choices are described in the following proposition and illustrated in Figure 3.

Proposition 3 (Lender-Optimal Transparency Regime)**.** *Let*

$$\pi_L(b,\phi) \triangleq \min\left\{\widehat{\pi}(b,\phi), \widehat{\pi}(X)\right\} \in (0, \widehat{\pi}(X)]\,, \tag{14}$$

where $\widehat{\pi}(X) \triangleq \left(\frac{1}{X}/(1-\frac{1}{X})\right)^2$ *while* $\widehat{\pi}(b,\phi)$ *is defined in Equation (44) and is increasing in* b *and* ϕ*. If the fraction of non-strategic borrowers is sufficiently high, i.e.* $\pi \geq \pi_L(b,\phi)$ *(resp. sufficiently low, i.e.* $\pi < \pi_L(b,\phi)$*), an opaque (resp. transparent) algorithm maximizes the lender's profits.*

Proof. See the Appendix. ∎

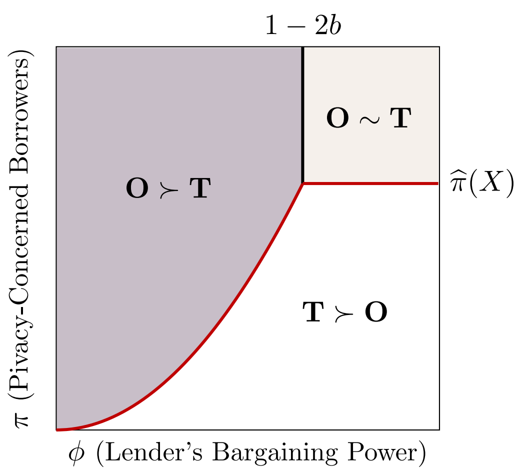


Figure 3: **Lender-Optimal Transparency Regime.** The figure depicts the lender's preferences between a transparent (**T**) and an opaque (**O**) algorithm as defined in Definition 2. The symbol $\succ$ denotes a strict preference, while $\sim$ denotes indifference. The red line represents $\pi_L(b,\phi)$ as defined in Equation (14).

The lender's optimal transparency regime maximizes the efficacy of the algorithm in rationing credit, foregoing the potential benefits of price discrimination. This happens because the Blackwell (1951) value of information is nil above the rationing threshold, as the lender is risk-neutral in this region. While detailed credit quality information can reduce the variance of the lender's returns from credit-worthy borrowers, it does not affect their expected value, since potential mispricings offset each other. Therefore, from an ex-ante perspective, the lender gains no additional profits from acquiring more granular information than what is necessary for optimal credit rationing. It follows that the algorithm's transparency regime is irrelevant in the region where the algorithm prescribes no credit rationing, i.e. when $\lambda \in (\underline{\lambda}, \overline{\lambda})$.

Building on this result, we can examine credit misallocations resulting from gaming when $\lambda \geq \overline{\lambda}$ (the analysis is symmetric when $\lambda \leq \underline{\lambda}$). Under transparency, when $\lambda \in \left[\overline{\lambda}, \widehat{\lambda}(\gamma(\pi))\right)$, strategic high-risk borrowers (with $z \in [0, r(\lambda)]$) evade credit rationing by withholding bad news. These borrowers represent a negative NPV investment for the lender and thus induce a loss in expectation. The magnitude of this loss increases with λ, as a greater number of high-risk borrowers slip through the lender's screening process and obtain credit. In addition, some marginally credit-worthy borrowers (with $z \in (r(\lambda, \gamma(\pi)))$ receive more favorable interest rates than what their actual credit risk would suggest. Because of risk neutrality, this second group of borrowers does not impact the lender's expected profits. For higher correlation levels, $\lambda \in \left[\widehat{\lambda}(\gamma(\pi)), 1\right]$, gaming does not occur. Only observationally credit-worthy borrowers receive credit, while non-disclosing borrowers are credit rationed. Within the group of strategic borrowers, the lender achieves profits comparable to those with full observability of data. However, the lender suboptimally refuses credit to non-strategic borrowers who are on average credit-worthy.

Overall, the profits the lender yields employing a transparent algorithm weakly decrease with the number of behavioral borrowers π. This happens for two reasons. First, as π increases, gaming behavior spreads across a wider range of allocation rules, exacerbating misallocations and resultant losses. Second, in the region where gaming is absent due to effective unraveling forces, an increasing number of non-strategic credit-worthy borrowers are erroneously denied credit. At a certain point, when $\pi \geq \widehat{\pi}(X) \in (0,1)$ i.e. $\widehat{\lambda}(\gamma(\pi)) \geq 1$, gaming extends throughout the entire range of λ and the lender's profits equal the no-information level.

Consider now credit misallocations resulting from an opaque algorithm. When the lender's bargaining power is sufficiently high (i.e. $\phi \geq 1-2b$), causing borrowers' hedging motives to be strong enough to prevent any data sharing in equilibrium, the lender is unable to perform either credit rationing or price discrimination, regardless of the true predictive power of the data. Consequently, the lender earns the lowest possible profits, equivalent to the no-information profits. Instead, when borrowers have relatively more bargaining power (i.e. $\phi < 1-2b$), they share extreme evidence hoping for favorable interest rates but may be excluded from credit ex-post. When correlation is mild, i.e. when $\lambda \in \left[\overline{\lambda}, \widehat{\lambda}(\eta(b, \phi))\right]$ or equivalently $r(\lambda) \leq \eta(b, \phi)$, high-risk borrowers with $z \leq r(\lambda) \leq \eta(b,\phi)$ share data, hoping to be classified as low risks, instead the algorithm uses the disclosed information to ration credit. Hence, only observationally credit-worthy borrowers obtain funding and the lender achieves profits comparable to those with full observability of data. When correlation is higher, $\lambda \in \big(\widehat{\lambda}(\eta(b,\phi)), 1\big]$ or equivalently $\eta(b,\phi) < r(\lambda)$, rationing occurs ex-post but is only partially optimal. In fact, hedging behavior precludes the lender from rationing credit to those high risks that withhold information, i.e. those with $z \in [\eta(b,\phi), r(\lambda)]$. Finally, non-strategic borrowers always obtain credit and this is optimal as they are on average creditworthy.

In summary, the lender's profits under opacity weakly decrease with the lender's bargaining power ϕ and the borrower's private benefit from getting credit, b. This is intuitive as these variables reinforce the borrower's hedging behavior and ex-post credit rationing occurs for a narrower range of allocation rules.

Transparency regimes can now be easily compared. When borrower's hedging motives are strong and gaming ability is high (i.e., $\phi \geq 1-2b$ and $\pi \geq \widehat{\pi}(X)$), the lender is unable to implement credit rationing, regardless of the chosen transparency regime. As a result, the lender is indifferent between the two regimes. For the remaining parameter values, the lender faces a trade-off. When gaming ability outweighs hedging motives, an opaque algorithm maximizes the lender's profits by extracting more information, thus enabling credit rationing for those high-risk borrowers who would otherwise game a transparent system. Conversely, when gaming ability is weaker than hedging motives, transparency becomes optimal by leveraging unraveling forces while eliminating hedging behavior. The threshold $\widehat{\pi}(b,\phi) \in \widehat{\pi}(X)\big)$ at which the lender switches from transparency to opacity is increasing in the lender's bargaining power ϕ and the borrower's private benefit b: as hedging motives strengthen, transparency becomes optimal for a broader range of parameters. When the borrower's private benefit exceeds $\frac{1}{2}$, hedging motives are so strong that an opaque algorithm results in no data-sharing irrespectively of the lender's bargaining power. As a consequence, transparency benefits the lender by enhancing data extraction and the algorithm's credit rationing capabilities.

6.2 Social Welfare

We now examine the socially optimal transparency regime — the one that maximizes egalitarian social welfare (or total surplus) — and show that the lender's transparency choices are often socially inefficient.

It is useful to first consider a benchmark in which the social planner can segment the data available to the lender, who then optimally allocates credit based on this segmentation (the approach is similar to Bergemann, Brooks and Morris (2015)). In this benchmark, the planner observes the lender's statistical technology, λ, and commits to a segmentation of observable credit risk, i.e. a monotone partition of the borrower's characteristics. Specifically, a credit risk segmentation is a partition $Z = \{z_0, z_1, \ldots, z_{n-1}, z_n\}$ of the borrower's data space $[0,1]$

with $0 = z_0 < z_1 < \cdots < z_{n-1} < z_n = 1$, where the lender only observes in which risk bucket r_i a borrower belongs, where $r_i \triangleq \{z \in [z_{i-1}, z_i)\}$ for $i \in \{1, \ldots, n\}$.

Every risk bucket r_i induces a lender-optimal allocation, that is a credit provision decision $a_\lambda(r_i)$ and an interest rate $x_\lambda(r_i)$, and thus can be seen as a recommendation about credit allocations from the planner to the lender. Hence, the welfare-optimal credit risk segmentation — that is the set of all these recommendations — has to i) maximize social welfare, and ii) be incentive-compatible (or obedient) for the lender.

Lemma 2 (Welfare-Optimal Credit Risk Segmentation)**.** *The welfare-optimal credit risk segmentation contains at most two risk buckets ($n \leq 1$) with cutoff*

$$z_1 = r^\star(\lambda, b) \triangleq \begin{cases} \max\left\{r(\lambda) - \frac{b}{\lambda X}, 0\right\} & \text{if } \lambda \in [0,1], \\ \min\left\{r(\lambda) - \frac{b}{\lambda X}, 1\right\} & \text{if } \lambda \in [-1,0), \end{cases} \tag{15}$$

where $r(\lambda)$ is defined in Equation (8).

Proof. See the Appendix. ■

The lemma shows that the market solution entails excessive rationing from a social perspective and the misalignment of incentives between the lender and the planner increases with the borrower's private benefit, b. To see this, note that the credit provision induced by the welfare-optimal segmentation (as a function of the underlying characteristic z) is

$$\ell^\star_\lambda(z) = \begin{cases} \mathbb{1}\left\{z > r(\lambda) - \frac{b}{\lambda X}\right\} & \text{if} \quad \lambda \in \left[\overline{\lambda}(b), 1\right] \\ 1 & \text{if} \quad \lambda \in \left(\underline{\lambda}(b), \overline{\lambda}(b)\right) \\ \mathbb{1}\left\{z < r(\lambda) - \frac{b}{\lambda X}\right\} & \text{if} \quad \lambda \in \left[-1, \underline{\lambda}(b)\right], \end{cases} \tag{16}$$

where $\overline{\lambda}(b) = \overline{\lambda} + \frac{2b}{X}$ and $\underline{\lambda}(b) = -\overline{\lambda}(b)$. Comparing this with the lender-optimal credit allocation in Equation (7) we find that the market solution results in insufficient provision of credit, as $\ell_\lambda(z) \leq \ell^\star_\lambda(z)$. Specifically, when $\lambda \geq \overline{\lambda}(b)$, there exists a subset of borrowers with $z \in (r(\lambda) - \frac{b}{\lambda X}, r(\lambda)]$ that would be rationed by the lender despite having a positive surplus project. The size of this set increases with b, as a higher private benefit for the borrower increases the overall surplus without changing the lender's profits. Nevertheless, in this region, the planner would deny credit to certain high-risk borrowers seeking to fund negative surplus projects. When $\lambda \in [\overline{\lambda}, \overline{\lambda}(b))$, the planner would extend credit to all borrowers, while the lender would still ration some with sufficiently low realizations $z \leq r(\lambda)$. The range of allocation rules for which this happens also increases with b. Only when b tends to 0, the lender implements the socially efficient level of credit provision.

Note also that the welfare-optimal segmentation entails at most two coarse risk buckets. This happens for two reasons. First, interest rates determine how the surplus is divided between the lender and the borrower, without altering the total surplus created. Therefore, the planner's only concern is to ensure welfare-optimal credit provision. Since credit provision is a binary decision, the segmentation contains at most two buckets as all the potential sub-buckets inducing the same decision can be pooled in a unique category. This is essentially due to the revelation principle, for which messages (risk buckets) can be thought of as incentive-compatible action recommendations. Second, the risk buckets have to be sufficiently coarse to be incentive-compatible. All the borrowers for which there is a misalignment of incentives between the lender and the planner have to be lumped in a unique risk category that the lender is willing to fund. For this to occur, the bucket has to contain a sufficient number of low-risk borrowers - together with high-risk - and thus has to be sufficiently coarse.

The following proposition and Figure 4 describe the transparency regime that maximizes social welfare.

Proposition 4 (Welfare-Optimal Transparency Regime)**.** *Let*

$$\pi_W(b, \phi) \triangleq \max\left\{0, \min\left\{\pi^{\circ}(b, \phi), \pi^{\circ}(b)\right\}\right\} < \pi_L(b, \phi), \tag{17}$$

where $\pi^{\circ}(b, \phi)$ *and* $\pi^{\circ}(b)$ *are defined in Equation (62) and Equation (63), respectively, and* $\min\left\{\pi^{\circ}(b, \phi), \pi^{\circ}(b)\right\}$ *is decreasing in* b *and weakly increasing in* ϕ*. If the fraction of non-strategic borrowers is sufficiently high, i.e.* $\pi \geq \pi_W(b, \phi)$ *(resp. sufficiently low, i.e.* $\pi < \pi_W(b, \phi)$*), an opaque (resp. transparent) algorithm maximizes social welfare.*

Proof. See the Appendix. ■

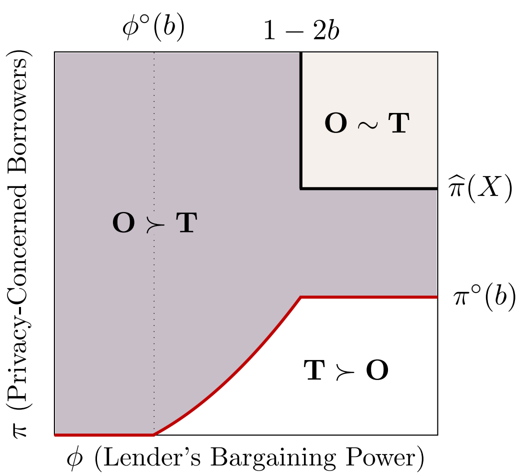


Figure 4: **Welfare-Optimal Transparency Regime.** The figure depicts the social planner's preferences between a transparent (**T**) and an opaque (**O**) algorithm as defined in Definition 2. The symbol $\succ$ denotes a strict preference, while $\sim$ denotes indifference. The red line represents $\pi_W(b, \phi)$ as defined in Equation (17), $\phi^{\circ}(b)$ is defined in Equation (56), while $\widehat{\pi}(X) \triangleq \left(\frac{1}{X}/(1 - \frac{1}{X})\right)^2$.

The welfare-optimal transparency regime depends on the borrower's private benefit through two key channels. First, as established in Lemma 2, the private benefit b creates a gap between socially efficient and lender-optimal credit provision. Second, in combination with the lender's bargaining power ϕ, the private benefit reinforces the borrower's hedging motives under an opaque algorithm. This, in turn, impacts the extent of information available to the lender and the resulting — potentially inefficient — credit rationing decisions.

When credit provision is socially efficient (i.e., when b approaches 0), the lender's transparency choices also are. In this case, an opaque algorithm is efficient (and optimal) for two reasons. First, opacity leaves to non-strategic credit-worthy borrowers the possibility to withhold data without being denied credit. Second, it maximizes data extraction from strategic borrowers — since hedging motives are absent and every credit-unworthy borrower shares data — and the disclosed information is used efficiently by the lender. On the other hand, under transparency, some negative-surplus projects would be funded because of gaming, while non-strategic borrowers would occasionally face rationing. It follows that no matter the lender's bargaining power and the mass of non-strategic borrowers, opacity is welfare-improving.

However, the lender's transparency decisions are often socially inefficient when preferences for credit inclusion deviate, even slightly, from those of the social planner. Moreover, the welfare-optimal regime favors more opacity than the market-driven solution, i.e., $\pi_W(b, \phi) < \pi_L(b, \phi)$. The social benefit of opacity arises because it eliminates the stigma typically associated with no disclosure in a transparent regime, thereby enabling privacy-concerned borrowers to fund positive-surplus projects. This benefit is only partially internalized by the lender, who captures just a fraction of the surplus generated by these projects. As a result, transparency in the credit market is excessive. This highlights a social benefit of opacity that is often overlooked by advocates of algorithmic transparency, who usually focus on equity, fairness, and truth as the primary social objectives. When the goal is improving credit market allocative efficiency, algorithmic opacity can perform surprisingly well and may promote important values such as financial inclusion and borrowers' privacy. Section 8 provides further insights into the policy implications of these findings.

To delve deeper into this mechanism, consider the case where the misalignment of preferences over credit inclusion is moderate, i.e. $b \in (0, b^{\circ}(X))$ where $b^{\circ}(X) \in (\frac{1}{4}, \frac{1}{3})$, meaning that data collection holds some social value by excluding strategic borrowers with negative-surplus projects. For simplicity, suppose also that opacity results in full hedging and no credit rationing ($\phi \geq 1 - 2b$) while transparency induces some credit rationing ($\pi < \widehat{\pi}(X)$).[27] In this scenario, the planner faces a trade-off. On one hand, transparency maximizes the lender's ability to extract information from strategic borrowers, allowing the exclusion of negative surplus projects. On the other hand, transparency comes with a stigma around data-withholding and leads to

[27]Similar effects are at play when both regimes induce some credit rationing ($\phi < 1 - 2b$ and $\pi < \big(\frac{1}{X}/(1 - \frac{1}{X})\big)^2$).

the exclusion of those privacy-concerned but creditworthy borrowers, who would secure credit under opacity. Transparency is socially efficient when the first effect dominates, that is when strategic borrowers are numerous enough, i.e. when $\pi < \pi^{\circ}(b)$. When $\pi \geq \pi^{\circ}(b)$, opacity is socially efficient. In this region, algorithms are overly transparent from a social standpoint, since the lender opts for transparency despite opacity being the welfare-maximizing choice (cfr. Figure 3). The inefficiency arises because the lender is more inclined to ration credit compared to the planner, and thus only partially internalizes the benefits of expanded credit access that comes with opacity.

As the misalignment over credit provision preferences widens, the lender's decisions become increasingly inefficient. When b exceeds the threshold $b^{\circ}(X)$, the cutoff $\pi^{\circ}(b)$ approaches 0 and opacity unambiguously maximizes welfare. Beyond this point, projects that would be rationed under transparency generate a positive social surplus from an ex-ante perspective so transparency has no social benefit. When $b > \frac{1}{2}$, the misalignment between the lender and the planner reaches its peak, leading to complete disagreement on the appropriate transparency regime to implement. Beyond this point, data extraction destroys surplus and the welfare optimal regime is the one that minimizes data-sharing, that is opacity.

6.3 Borrower's Surplus and Redistributive Effects

We now analyze borrowers' surplus across different transparency regimes, distinguishing between privacy-concerned and strategic borrowers. Strategic borrowers' surplus depends on their observable credit risk, as this influences their allocations trough their data-sharing decisions. In contrast, the surplus of privacy-concerned borrowers is independent of data as they consistently withhold information.

Proposition 5 (Redistributive Effects)**.** *The surplus of privacy-concerned borrowers is strictly lower when the lender employs a transparent algorithm. The surplus of strategic borrowers is strictly higher when the lender employs a transparent algorithm only if their data realization is extreme, i.e.* $z \in [0, z^{\star}(\pi)) \cup (1 - z^{\star}(\pi), 1]$ *where* $z^{\star}(\pi) \in [0, \frac{1}{2})$ *is weakly increasing in* π *and defined in Equation (85).*

Proof. See the Appendix. ■

Privacy-concerned borrowers are better off under opacity. As noted in Corollary 1, a transparent algorithm impose a negative externality on them through the stigma associated with no-disclosure. Under transparency, they face higher interest rates compared to opacity and may even be excluded from credit.

Strategic borrowers favor transparency when their data is extreme and privacy-concerned borrowers sufficiently numerous. This is because transparency provides borrowers the option to withhold bad news while sharing good news, and this option value is the greatest for extreme

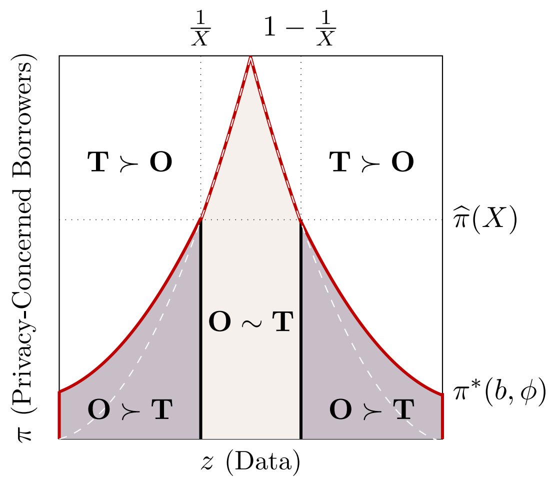


Figure 5: **Transparency Regime Preferred by Strategic Borrowers.** The figure depicts the strategic borrowers' preferences between a transparent (**T**) and an opaque (**O**) algorithm as defined in Definition 2, when $\phi \geq 1 - 2b$. The symbol $\succ$ denotes a strict preference, while $\sim$ denotes indifference. The red lines represent $z^\star(\pi)$ and $1 - z^\star(\pi)$ where $z^\star(\pi)$ is defined in Equation (85), while $\widehat{\pi}(X)$ and $\pi^*(b, \phi)$ are defined in Equations (72) and (79), respectively. The white dashed lines represent $\gamma(\pi)$ and $1 - \gamma(\pi)$ where $\gamma(\pi)$ is defined in Equation (12).

borrowers if bad news can be easily concealed. When they share good news, they stand to gain the most, as their data provides a strong, highly indicative signal of credit quality due to its extreme nature. When they withhold bad news (and still obtain credit), they achieve the greatest improvement in contractual terms by being grouped with borrowers of significantly higher credit quality than their actual, observable risk level.

To elaborate on this further, consider the case where opacity leads to complete data-withholding — i.e. when the lender's bargaining power is sufficiently high, $\phi \geq 1 - 2b$ — and the lender offers credit to all borrowers at a single interest rate.[28] This situation is reported in Figure 5. When privacy-concerned borrowers are sufficiently numerous, i.e. $\pi \geq \widehat{\pi}(X)$, the stigma around data-withholding is mild and every borrower gets credit even under transparency. Central borrowers — those with $z \in (z^\star(\pi), 1 - z^\star(\pi)) = (\gamma(\pi), 1 - \gamma(\pi))$ — always share information, and because data is treated symmetrically across allocation rules, the average interest rate they receive is the same under both transparent and opaque regimes, leaving them indifferent between the two. In contrast, by withholding data, extreme borrowers obtain rates that are significantly lower than what their observable risk would suggest, resulting in lower average rates under transparency. The improvement in interest rates grows as the gap between true and inferred credit risk widens, hence when data is more extreme and privacy-concerned borrowers more numerous. When the stigma of no-disclosure is moderate, i.e. $\pi < [\pi^*(b, \phi), \widehat{\pi}(X))$, withholding data may lead to credit denial. The option value of

[28] Similar effects are at play when $\phi < 1 - 2b$. See the proof of Proposition 5 for a formal treatment.

data withholding reduces and central borrowers — those with $z \in (z^\star(\pi), 1 - z^\star(\pi))$ with $z^\star(\pi) < \gamma(\pi)$ — weakly prefer the security of credit provided by opacity. When unraveling forces increases even further, i.e. $\pi < \pi^\star(b, \phi)$, every borrower prefers the safe hedge of opacity.

Aggregating over credit risk and privacy types we obtain the aggregate borrower's surplus, and the following result.

Proposition 6 (Borrower-Optimal Transparency Regime)**.** *An opaque algorithm maximizes ex-ante borrower's surplus.*

Proof. See the Appendix. ∎

Although strategic borrowers may sometime prefer a transparent algorithm, the overall borrower's surplus is higher when the lender employs an opaque algorithm. This happens because strategic borrowers favor transparency precisely when they are few, and thus have a smaller weight in aggregate surplus. Ultimately, the security of credit provision offered by opacity outweighs the advantages of selective disclosure enabled by transparency, resulting in an increase in aggregate borrower surplus.

7 Empirical Implications

The above analysis offers a number of testable implications, with some findings supported by recent studies and others presenting opportunities for further research.

Recent papers show that information has distributional effects in credit markets (see, e.g., Liberman et al. (2019), Dobbie et al. (2020), Nelson (2023), and Jansen et al. (2024)). My paper suggests that this holds even when borrowers have control over information sharing - regardless of the algorithm's transparency regime - and data sharing may reduce credit provision compared to a scenario where no information is shared.[29] While some studies find that Open Banking enhances credit provision (see, e.g., Nam (2024)), others observe that the policy creates both winners and losers, in line with my findings. For instance, Doerr et al. (2023) show that the California Consumer Privacy Act — giving borrowers more control over their data — led to increased data-sharing with FinTech lenders, which in turn resulted in higher denial rates and greater dispersion in interest rates, despite an overall reduction in the average interest rate. Babina et al. (2023) find that borrowers that were denied credit by their relationship banks did not see improved access to credit after the implementation of the Commercial Credit Data Sharing in the UK, as the shared data marked them as high credit risk and reduced their ability to establish new lending relationships. Similarly, Rishabh (2024) find that payment data, central to Open Banking initiatives, benefits most borrowers but disadvantages high credit risks.

[29] Note that in my model, by Assumption 1, the lender extends credit to the entire pool of borrowers in the absence of information.

In line with my differentiation between non-strategic and strategic agents, Lin (2022) empirically separates two components of consumers' privacy preferences: an intrinsic component (an exogenous taste for privacy) and an instrumental component (an endogenous economic loss from revealing private information). Critically, the study is conducted in a controlled experimental setting, where "how the instrumental incentive depends on a consumer's type is straightforward and explicitly explained" to participants, akin to a transparent regime in my model. The paper's empirical findings are consistent with my model. In particular, consumers self select into data sharing depending on the respective magnitude of the two components and exhibit gaming type of behavior: high-type consumers are more willing to share data, while low types tend to withhold data, as the receiver's equilibrium inference following no disclosure is attenuated by intrinsic motives. The study could potentially be extended to an opaque regime to investigate whether the receiver's equilibrium inference is less averse compared to a transparent setting and whether senders adopt hedging behaviors in response to opacity.

Some research has shown that intrinsic motives for privacy are particularly pronounced in credit markets. Tang (2019) studies loan applications to a Chinese Fintech lender where the disclosure of personal data is a pre-condition for loans, thus allowing her to isolate borrowers' intrinsic preferences for privacy.[30] The paper shows that stricter disclosure requirements reduce loan application completion rates, suggesting that borrowers' intrinsic privacy concerns are strong enough to outweigh the potential benefits of obtaining a loan.[31] This finding is in line with my model's assumption that privacy-concerned borrowers consistently withhold information. It also suggests that unraveling forces may drive such borrowers out of the market in a transparent regime. Whether this finding is supported by the data remains an open question for future empirical investigation.

My analysis provides fresh insights into earlier studies on borrowers' data-sharing behavior. Notably, Nam (2024) examines borrowers' voluntary data-sharing decisions with a major German FinTech lender. The study takes place under a relatively opaque regime, as the platform discloses that shared information ("behavioral data, web data, and experience data") could potentially lead to credit denial or higher interest rates, but offers little explanation of the underlying decision logic. First, the study finds that the average rate of data sharing was quite low, at just 8% across the entire sample period. Second, it finds no evidence of lender's equilibrium inference in response to data withholding, as an increase in data sharing results in a negligible negative impact on loan approval for those who choose not to disclose. While these findings are inconsistent with the standard unraveling logic (e.g., Grossman (1981) and

[30]The paper shows that the disclosure group is not statistically less risky than the no-disclosure group, supporting the causal impact of intrinsic (rather than instrumental) motive for privacy on application withdrawal.

[31]As the author acknowledges, all Chinese platforms commonly ask for non-standard personal information from borrowers, as there is no official credit score system and there are limited resources available to verify borrower credentials and documents. Hence, Chinese borrowers have limited ability to substitute for other, less privacy-intrusive platforms.

Milgrom (1981)), they align with the equilibrium proposed in Proposition 2, where borrowers hedge against opacity by withholding information, and the lender does not penalize the absence of disclosure.

Recent papers provide suggestive evidence for the primary mechanism in my model, that is firms leverage their transparency regime as a strategy to maximize data collection (Proposition 3). First, in light of my theory, firms with greater bargaining power should be more transparent to mitigate borrowers' hedging motives. Relatedly, Ramadorai, Uettwiller and Walther (2021) show that larger firms are more likely to have a privacy policy, to display it visibly, to write longer and more sophisticated policies (conditional on having one). These firms also engage in more extensive data extraction, utilizing a greater number of cookies to track consumer behavior. These findings support the notion that transparency serves as a strategic tool to reassure users and enhance data extraction. Second, my model suggests that firms facing more privacy-concerned borrowers should be more opaque. Relatedly, Bian, Ma and Tang (2021) show that mandated disclosure of privacy labels in the AppStore negatively impacted firms' profits, with this effect being more pronounced in countries with stronger privacy concerns. In other words, the lack of transparency in firm data collection practices led consumers to share excessive data. This suggests that in markets with stronger privacy concerns, firms have greater incentives to be opaque. Although these findings center on input data transparency rather than model transparency, they indicate that firms strategically optimize transparency to maximize data collection.

My model could inform further empirical investigation on this mechanism. According to it, financial institutions with greater bargaining power should reveal finer details about the inner workings of their algorithm — not only about the input data they use — to mitigate mistrust and foster information sharing. Conversely, financial institution facing borrowers with stronger privacy concerns should reveal less about their algorithm's operating logic, to prevent gaming type of behavior. In addition, my theory offers a number of testable implications on the link between algorithmic transparency, data sharing and credit market outcomes not yet explored by the empirical literature. In Appendix B, I present closed-form computations of key observables in my model (such as amount of data shared, levels of credit provision, and interest rates) to guide hypothesis testing.

The above discussion raises the question of how to measure primitive variables in my model, particularly an algorithm's transparency regime. Computer scientists have developed quantitative indicators of AI transparency based on publicly available information. For example, the Foundation Model Transparency Index (see Bommasani et al. (2023)) codifies AI transparency by incorporating various aspects, including data sources, model development, and decision-making processes, and reveals significant variation across firms. A similar index could be tailored specifically for credit scoring models. More qualitative measures of transparency could include explainability scores, which assess how well the algorithm's decisions can be understood, combined with disclosure levels, that is the extent to which information about the algorithm — including data used, methodologies, and underlying logic — is made publicly available on a lender's website. The staggered implementation of the AI Act (see Section 9) could introduce an additional source of variation. The second variable of interest is borrowers' privacy concerns, for which the literature provides various proxies. Research has demonstrated that consumers' privacy concerns in online environments correlate with several factors, including cultural influences (such as trust in institutions), demographic variables (like age, education level, and income), trust in the firm (encompassing reputation, credibility, and brand recognition), and the nature of the data shared (including breadth, depth, and sensitivity).[32] In the credit market, privacy concerns vary across countries, types of data, demographic and income groups (see Tang (2019), Babina et al. (2023), Doerr et al. (2023) and Nam (2024)). Lastly, a lender's bargaining power can be effectively measured using traditional indicators of market power and concentration, such as market share or the Herfindahl-Hirschman Index.

8 Policy Discussion

The EU Artificial Intelligence (AI) Act leads global efforts to regulate the development, deployment, and use of artificial intelligence technologies. The regulation entered into force on August 1$^{\text{st}}$ 2024 and shall apply from August 2$^{\text{nd}}$ 2026 (see Article 113). According to the Act, credit scoring and credit underwriting models are classified as high-risk AI systems (see Article 6(2) and Annex III, 5(b)). As such, they are subject to various transparency requirements aimed at ensuring accountability and clarity in their operations. In addition to mandating transparency towards deployers and the Commission,[33] the Act imposes both ex-post (Article 86) and ex-ante (Article 71) disclosure requirements towards the general public. Article 71 specifies that the Commission will collect information from developers to set up and maintain a database containing, among other things, a "description of the information used by the system (data, inputs) and its operating logic" (see Annex VIII, Section A(6)). This information should "accessible and publicly available in a user-friendly manner".

My model suggests that the transparency requirements of the AI Act may impair the credit market's allocative efficiency in some cases, enabling gaming behavior and disproportionately harming privacy-concerned borrowers, somehow at odds with the regulator's stated

[32]See Acquisti, Taylor, and Wagman (2016) for a general discussion and Morey, Forbath, and Schoop (2015), or Armantier et al. (2021), or Prince and Wallsten (2022) for surveys.

[33]Credit scoring models should be explainable, that is they "should be designed and developed in such a way as to ensure that their operation is sufficiently transparent to enable deployers to interpret a system's output and use it appropriately" (see Article 13). Moreover, models should be auditable, as the Commission can require firms to provide technical documentation including a general description of the AI model - including the architecture, the number of parameters, the modality and format of inputs and outputs - and a detailed description of the elements of the model - including the design specifications, training process, training methodologies, training techniques, design choices, assumptions made, what the model is designed to optimize for, training, testing and validation datasets, their scope and main characteristics (see Article 91, Article 53, and Annex XI).

commitment to privacy in frameworks like the GDPR. This finding enriches the current policy debate surrounding the AI Act by offering some high-level insights. First, it suggests that the allocative efficiency of the impacted markets should not be overlooked. Currently, the regulation's core guiding principles — human oversight, accountability, safety, fairness and non-discrimination — focus primarily on ethical and safety considerations, ignoring implications for market efficiency. Second, an industry-based approach should complement the existing risk-based, technology-neutral framework. In the finance industry, lenders' preferences for financial inclusion often conflict with social efficiency, leading to inefficient transparency choices by financial institutions. The gap between social and private goals may be narrower in other industries so an industry-based approach could help regulators better allocate limited resources for more effective regulation. Third, regulations on algorithmic transparency should be implemented in tandem with data privacy regulations, such as the GDPR. Algorithmic transparency directly influences individuals' privacy choices, making it essential to coordinate both areas to ensure comprehensive protection for users.

My model also offers concrete recommendations for regulators, helping them establish clearer guidance on compliance expectations within the financial services industry.

- **Selective Transparency** Algorithmic transparency should only be mandated when a lender's market power is significant enough to prompt borrowers' hedging behavior against opacity (e.g. for BigTechs) and for information for which borrowers have minimal privacy concerns (e.g. basic demographic information). When a lender's market power is limited, or privacy concerns are heightened, preserving opacity enhances market efficiency.

- **Operating Logic** Regulators should clarify what is meant by an algorithm's *operating logic* in Article 71. My model suggests that disclosing the *direction* in which data influences predictions is an effective and practical approach to transparency. It can be shown that revealing only the sign of λ to borrowers results in the same credit allocation as disclosing its exact value. Intuitively, simply knowing whether higher values of a variable are seen as good or bad news is sufficient to create a stigma around non-disclosure, triggering unraveling forces. Providing more granular information about λ is sometimes impractical, especially with complex machine learning models, while direction-based disclosure is feasible thanks to recent explainable AI techniques. This laxer requirement would enable financial institutions to continue using high-performing algorithms without compromising their screening ability. On the other hand, direction-based disclosure is essential for achieving transparency. Leaving uncertainty about whether data signals good or bad news prevents unraveling forces and activates hedging behavior, ultimately preserving opacity.

- **Access to the Database** Regulators should require financial institutions to provide borrowers direct access to the database mentioned in Article 71 during loan applications, or ideally, have lenders disclose the necessary information on their websites. In light of my model, transparency can be more effectively achieved when borrowers understand the algorithm's operating logic at the time of the data-sharing decision, and when lenders are aware that borrowers possess this information. Establishing the database without ensuring borrowers are informed about it could lead to situations where only some borrowers are aware of its existence, or where only a few are willing to bear the costs to access the information. This would potentially create unequal access to the database, leading some borrowers to operate under transparency while others under opacity and complicating the lender's equilibrium inference.

- **Price for Data** In addition to promoting transparency, regulators should encourage financial institutions to implement measures for screening borrowers' privacy types. This perspective aligns with the ongoing debate about introducing a price for personal data (see, e.g., Acquisti, Taylor and Wagman (2016)). In my model, within the transparency regime, identifying whether a borrower is privacy-concerned or strategic would enable the lender to extend credit to creditworthy privacy-concerned borrowers who might otherwise be excluded from the market due to negative externalities (see Corollary 1). On the other hand, unraveling forces would induce optimal, but sometimes inefficient, credit rationing of strategic users. Hence, screening enhances lenders' profits (net of agency rents) and also improves social welfare when the first effect dominates (i.e. when privacy-concerned borrowers are sufficiently numerous). Some financial institutions are already taking steps in this direction. For instance, Auxmoney, a prominent German FinTech lender, features what appears to be a menu of contractual terms on its website designed to encourage disclosure from borrowers who are less concerned about privacy. The lender showcases that sharing personal data with the platform may result in an *average* discount of €390 on a €5,000 loan, though interest rates could increase or applications might still be declined (see Nam (2024)).

9 Conclusion

This paper studies whether credit risk algorithms should be transparent or opaque, considering their impact on borrowers' data-sharing decisions. Transparency exposes the lender's model to gaming through strategic withholding of unfavorable information. Opacity mitigates gaming but leads borrowers to withhold information as a hedge against the unpredictability of the black box. The lender's transparency choices are aimed at data extraction and are influenced by the lender's bargaining power, which affects borrowers' hedging motives, as well as borrowers' privacy concerns that enhance their ability to game the system. These choices often lead to inefficiencies due to the lender's excessive inclination toward credit rationing.

Surprisingly, algorithmic opacity is often socially efficient because it removes the stigma typically associated with non-disclosure, thereby promoting credit access for privacy-concerned, creditworthy borrowers.

Several questions remain open for further research. First, the optimal transparency regime may lie in between full transparency and full opacity. Second, the analysis may be extended to data used in the monitoring process - rather than screening - or to more complex financial contracts. Third, competition may influence the lender's strategic choices of opacity. Lastly, algorithmic opacity could be examined in other financial contexts, such as algorithmic trading and portfolio management, as well as in other markets, including online ranking, targeted marketing, health care, demand forecasting, and fraud detection.

References

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

Agarwal, S., Alok, S., Ghosh, P. and Gupta, S., (2019), Financial Inclusion and Alternate Credit Scoring: Role of Big Data and Machine Learning in Fintech, Working Paper.

Ali, S. N., Lewis, G. and Vasserman, S., (2023), Voluntary Disclosure and Personalized Pricing, *Review of Economic Studies*, 90(2), pp. 538–571.

Armantier, O., Doerr, S., Frost, J., Fuster, A. and Shue, K., (2021), Whom Do Consumers Trust with their Data? US Survey Evidence, BIS Bulletins, 42, Bank for International Settlements.

Athey, S., Bryan, K. and Gans, J., (2020), The Allocation of Decision Authority to Human and Artificial Intelligence, *AEA Papers and Proceedings*, 110, pp. 80-84.

Babina, T., Bahaj, S., Buchak, G., De Marco, F., Foulis, A., Gornall, W., Mazzola, F. and Yu, T., (2023), Customer Data Access and Fintech Entry: Early Evidence from Open Banking, *Journal of Financial Economics*, forthcoming.

Ball, I., (2024), Scoring Strategic Agents, *American Economic Journal: Microeconomics*, forthcoming.

Bartlett, R., Morse, A., Stanton, R. and Wallace, N., (2022), Consumer-lending Discrimination in the FinTech Era, *Journal of Financial Economics*, 143(1), pp. 30-56.

Bazarbash, M., (2019), FinTech in Financial Inclusion: Machine Learning Applications in Assessing Credit Risk, IMF Working Paper No. 2019/109.

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), pp. 2845-2897.

Bergemann, D. and Bonatti, A., (2019), Markets for Information: An Introduction, *Annual Review of Economics*, 11, pp. 85-107.

Bergemann, D., Brooks, B. and Morris, S., (2015), The limits of price discrimination, *American Economic Review*, 105, pp. 921–957.

Bergemann, D. and Morris, S., (2019), Information Design: A Unified Perspective, *Journal of Economic Literature*, 57(1), pp. 44-95.

Besanko, D. and Thakor, A., (1987), Collateral and Rationing: Sorting Equilibria in Monopolistic and Competitive Credit Markets, *International Economic Review*, 28(3), pp. 671-89.

Bester, H., (1985), Screening vs. Rationing in Credit Markets with Imperfect Information, *American Economic Review*, 75(4), pp. 850-55.

Bian, B., Ma, X. and Tang, H., (2021), The Supply and Demand for Data Privacy: Evidence from Mobile Apps, Working Paper.

Binmore, K., (1987), "Perfect Equilibria in Bargaining Models", in *The Economics of Bargaining*, Basil Blackwell, Oxford.

Björkegren, D., Blumenstock, J. E. and Knight, S., (2021), Manipulation-Proof Machine Learning, Working Paper.

Blackwell, D., (1951), "Comparison of Experiments". In J. Neyman (Ed.), Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press.

Blattner, L., Nelson, S. and Spiess, J., (2024), Unpacking the Black Box: Regulating Algorithmic Decisions, Working Paper.

Blickle, K., He, Z., Huang, J. and Parlatore, C., (2024), Information-Based Pricing in Specialized Lending, NBER Working Paper No. w32155.

Bommasani, R., Klyman, K., Longpre, S., Kapoor, S., Maslej, N., Xiong, B., Zhang, D. and Liang, P., (2023), The Foundation Model Transparency Index, Working Paper.

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

Broecker T., (1990), Credit-Worthiness Tests and Interbank Competition, *Econometrica*, 58(2), pp. 429-452.

Cohn, J., Rajan, U. and Strobl, G., (2024), Credit Ratings: Strategic Issuer Disclosure and Optimal Screening, *Review of Finance*, forthcoming.

Cowgill, B. and Tucker, C., (2020), Algorithmic Fairness and Economics, Columbia Business School Research Paper.

Dobbie, W., Goldsmith-Pinkham, P., Mahoney, N. and Song, J., (2020), Bad Credit, No Problem? Credit and Labor Market Consequences of Bad Credit Reports, *The Journal of Finance*, 75(5), pp. 2377-2419.

Doerr, S., Gambacorta, L., Guiso, L. and Sanchez del Villar, M., (2023), Privacy Regulation and Fintech Lending, BIS Working Paper, No. 1103.

Dworczak, P. and Martini, G., (2019), The Simple Economics of Optimal Persuasion, *Journal of Political Economy*, 127(5), pp. 1993-2048.

Dye, R. A., (1985), Disclosure of Nonproprietary Information, *Journal of Accounting Research*, 23(1), pp. 123-145.

Ederer, F., Holden, R. and Meyer, M. A., (2018), Gaming and Strategic Opacity in Incentive Provision, *RAND Journal of Economics*, 49(4), pp. 819-854.

Ekmekci, M., Gorno, L. and Maestri, L., (2022), Learning from Manipulable Signals, *American Economic Review*, 112(12), pp. 3995-4040.

Fishman, M. J. and Hagerty, K. M., (2003), Mandatory Versus Voluntary Disclosure in Markets with Informed and Uninformed Customers, *The Journal of Law, Economics, and Organization*, 19(1), pp. 45-63.

Frankel, A. and Kartik, N., (2019), Muddled Information, *Journal of Political Economy*, 127(4), pp. 1739-1776.

Frankel, A. and Kartik, N., (2021), Improving Information from Manipulable Data, *Journal of the European Economic Association*, 20(1), pp. 79–115.

Freixas, X. and Laffont, J. J., (1990), "Optimal Banking Contracts", in *Essays in Honor of Edmond Malinvaud*, MIT Press, Vol. 2: Macroeconomics, pp. 33–61.

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

Gamba, A. and Hennessy, C., (2024), Manipulable Data, Goodhart's Law, and Credit Risk Prediction, Working Paper.

Gambacorta, L., Huang, Y., Li, Z., Qiu, H. and Chen, S., (2020), Data vs Collateral, BIS Working Paper, No. 881.

Glikson, E. and Woolley, A. W., (2020), Human trust in artificial intelligence: Review of empirical research, *Academy of Management Annals*, 14(2), pp. 627–660.

Goldfarb, A. and Tucker, C., (2012), Shifts in Privacy Concerns, *American Economic Review: Papers & Proceedings*, 102(3), pp. 349–353.

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

Harbaugh, R. and To, T., (2020), False Modesty: When Disclosing Good News Looks Bad, *Journal of Mathematical Economics*, 87, pp. 43-55.

Hardt, M., Megiddo, N., Papadimitriou, C. and Wooters, M., (2016), "Strategic Classification", in *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, pp. 111–122.

He, Z., Huang, J. and Parlatore, C., (2024), Information Span in Credit Market Competition, Working Paper.

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

Holmström, B. and Tirole, J., (1997), Financial Intermediation, Loanable Funds, and the Real Sector, *Quarterly Journal of Economics*, 112(3), pp. 663–691.

Huang, J., (2023), Fintech Expansion, Working Paper.

Iyer, R., Khwaja, A. I., Luttmer, E. F. P. and Shue, K., (2016), Screening Peers Softly: Inferring the Quality of Small Borrowers, *Management Science*, 62(6), pp. 1554–1577.

Jansen, M., Nagel, F., Yannelis, C. and Zhang, A. L., (2024), Data and Welfare in Credit Markets, Working Paper.

Jung, W.-O. and Kwon, Y. K., (1988), Disclosure when the Market is Unsure of Information Endowment of Managers, *Journal of Accounting Research*, 26(1), pp. 146-153.

Kamenica, E. and Gentzkow, M., (2011), Bayesian Persuasion, *American Economic Review*, 101(6), pp. 2590-2615.

Kleinberg, J., Ludwig, J., Mullainathan, S. and Rambachan, A., (2018), Algorithmic Fairness, *AEA Papers and Proceedings*, 108, pp. 22-27.

Korinek, A. and Balwit, A., (2022), Aligned with Whom? Direct and Social Goals for AI Systems, NBER Working Paper No. w30017.

Liang, A., Lu, J., Mu, X. and Okumura, K., (2024), Algorithm Design: A Fairness-Accuracy Frontier, *Journal of Political Economy*, forthcoming.

Liberman, A., Neilson, C., Opazo, L. and Zimmerman, S. D., (2019), The Equilibrium Effects of Information Deletion: Evidence from Consumer Credit Markets, Chicago Booth Research Paper No. 19-15, Fama-Miller Working Paper.

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

Leitner, Y. and Williams, B., (2023), Model Secrecy and Stress Tests, *Journal of Finance*, 78, pp. 1055-1095.

Lewis, T. and Sappington, D., (1994), Supplying Information to Facilitate Price Discrimination, *International Economic Review*, 35(2), pp. 309-327.

Milgrom, P. (1981), Good News and Bad News: Representation Theorems and Applications, *Bell Journal of Economics*, 12(2), pp. 380-391.

Molnar, C. (2023), Interpretable Machine Learning: a Guide for Making Black Box Models Explainable, 2nd ed., Leanpub.

Morey, T., Forbath, T., and Schoop, A. (2015), Customer Data: Designing for Transparency and Trust, *Harvard Business Review*, 93(5), pp. 96-105.

Nam, R. J. (2024), Open Banking and Customer Data Sharing: Implications for FinTech Borrowers, Working Paper.

Nelson, S. (2023), Private Information and Price Regulation in the US Credit Card Market, *Econometrica*, forthcoming.

Pasquale, F. (2015), The Black Box Society: the Secret Algorithms that Control Money and Information, Harvard University Press.

Perez-Richet, E. and Skreta, V. (2022), Test Design under Falsification, *Econometrica*, 90(3), pp. 1109-1142.

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

Quigley, D. and Walther, A. (2024), Inside and Outside Information, *Journal of Finance*, 79(4), pp. 2667-2714.

Rajan, R. (1992), Insiders and Outsiders: the Choice between Informed and Arm's-length Debt, *Journal of Finance*, 47(4), pp. 1367-400.

Ramadorai, T., Uettwiller, A., and Walther, A. (2021), Privacy Policies and Consumer Data Extraction: Evidence from U.S. Firms, Working Paper.

Rambachan, A., Kleinberg, J., Mullainathan, S., and Ludwig, J. (2021), An Economic Approach to Regulating Algorithms, NBER Working Paper Series, Working Paper 27111.

Rishabh, K. (2024), Beyond the Bureau: Interoperable Payment Data for Loan Screening and Monitoring, Working Paper.

Rubinstein, A. (1985), Perfect Equilibrium in a Bargaining Model, *Econometrica*, 50, pp. 97-110.

Sharpe, S. (1990), Asymmetric Information, Bank Lending, and Implicit Contracts: a Stylized Model of Customer Relationships, *Journal of Finance*, 45(4), pp. 1069-87.

Sun, J. (2021), Algorithmic Transparency, Working Paper.

Tang, H. (2019), The Value of Privacy: Evidence from Online Borrowers, Working Paper.

Verrecchia, R. E. (1983), Discretionary Disclosure, *Journal of Accounting and Economics*, 5(1), pp. 179-194.

Vives, X. (2019), Digital Disruption in Banking, *Annual Review of Financial Economics*, 11, pp. 243-272.

Wang, Q., Huang, Y., Jasin, S., and Singh, P. V. (2023), Algorithmic Transparency with Strategic Users, *Management Science*, 69(4), pp. 1935-2545.

Appendix

Proof of Lemma 1. We proceed in 2 steps: Step 1 derives the lender's estimate of the borrower's quality given data z, while Step 2 finds the optimal credit allocation.

Step 1. Given the DGP in Equation (1), the conditional pdf is

$$f_\lambda(\theta|z) = \begin{cases} \lambda\,\delta(\theta - z) + (1-\lambda) & \text{if } \lambda \geq 0 \\ |\lambda|\,\delta(1-\theta-z) + (1-|\lambda|) & \text{if } \lambda < 0, \end{cases}$$

where

$$\delta(t-z) \triangleq \begin{cases} \infty & \text{if } t-z=0 \\ 0 & \text{if } t-z \neq 0, \end{cases} \tag{18}$$

is the Dirac's delta function, satisfying the following property:

$$\int_{s-a}^{s+a} f(t)\,\delta(t-z)\,\mathrm{d}t = f(z) \quad \forall a > 0. \tag{19}$$

When $\lambda \geq 0$ we have

$$\begin{aligned} \mathbb{E}(\theta|z,\lambda) &= \int_0^1 \theta f_\lambda(\theta|z)\,\mathrm{d}\theta \\ &= \lambda \int_0^1 \theta\,\delta(\theta-z)\,\mathrm{d}\theta + (1-\lambda)\int_0^1 \theta\,\mathrm{d}\theta \\ &= \lambda z + (1-\lambda)\,\tfrac{1}{2}. \end{aligned}$$

When $\lambda < 0$, performing similar computations, we get

$$\begin{aligned} \mathbb{E}(\theta|z,\lambda) &= |\lambda|(1-z) + (1-|\lambda|)\,\tfrac{1}{2} \\ &= -\lambda\,(1-z) + (1+\lambda)\,\tfrac{1}{2} \\ &= \lambda\, z + (1-\lambda)\,\tfrac{1}{2}. \end{aligned}$$

Step 2. The first order conditions of the problem in Equation (5) are

$$\phi\Big[\mu_\lambda(z)x - 1\Big]^{\phi-1}\mu_\lambda(z)\Big[\mu_\lambda(z)\big(X-x\big)\Big]^{1-\phi} - \Big[\mu_\lambda(z)x-1\Big]^{\phi}(1-\phi)\Big[\mu_\lambda(z)\big(X-x\big)\Big]^{-\phi}\mu_\lambda(z) = 0.$$

Moving the second addend on the right-hand side and dividing both sides by

$$(1-\phi)\Big[\mu_\lambda(z)x-1\Big]^{\phi-1}\mu_\lambda(z)\Big[\mu_\lambda(z)\big(X-x\big)\Big]^{1-\phi}$$

we get

$$\frac{\phi}{1-\phi} = \frac{\mu_\lambda(z)x-1}{\mu_\lambda(z)\big(X-x\big)},$$

or equivalently

$$\phi = \frac{\mu_\lambda(z)x - 1}{\mu_\lambda(z)X - 1}.$$

Solving for x we get the optimal interest rate

$$x_\lambda(z) = \frac{1}{\mu_\lambda(z)} + \phi\left(X - \frac{1}{\mu_\lambda(z)}\right).$$

The lender provides credit only if her expected profits are strictly positive, i.e. only if

$$\mu_\lambda(z)\, x_\lambda(z) - 1 = \phi\Big(\mu_\lambda(z)X - 1\Big) > 0,$$

so the optimal credit rationing rule is

$$\ell_\lambda(z) = \mathbb{1}\left\{\mu_\lambda(z) > \frac{1}{X}\right\}.$$

For future reference, I compute the lender's expected profits, the borrower's expected profits and expected social surplus at the optimum (plugging the optimal allocation in Equation (3) and Equation (4) and taking expectation)

$$\begin{aligned} V_\lambda^L(z) &= \ell_\lambda(z)\Big[\phi\Big(\mu_\lambda(z)X - 1\Big)\Big], \\ V_\lambda^B(z) &= \ell_\lambda(z)\Big[(1-\phi)\Big(\mu_\lambda(z)X - 1\Big) + b\Big], \\ W_\lambda(z) &= \ell_\lambda(z)\Big[\mu_\lambda(z)X - 1\Big]. \end{aligned} \tag{20}$$

■

Lemma A.1 (Credit Allocation following Data Withholding)**.** *Let $Q \triangleq \{z \in [0,1] : m(z) = \varnothing\}$ be the set of borrowers that withhold data, the optimal credit allocation is*

$$\begin{aligned} x_\lambda(\varnothing) &= \frac{1}{\mu_\lambda(\varnothing)} + \phi\left(X - \frac{1}{\mu_\lambda(\varnothing)}\right), \\ \ell_\lambda(\varnothing) &= \mathbb{1}\left\{\mu_\lambda(\varnothing) > \frac{1}{X}\right\}. \end{aligned} \tag{21}$$

where

$$\begin{aligned} \mu_\lambda(\varnothing) &\triangleq \lambda\, z(\pi, Q) + (1-\lambda)\tfrac{1}{2}, \\ z(\pi, Q) &\triangleq \omega(\pi, Q)\tfrac{1}{2} + \big(1 - \omega(\pi, Q)\big)\mathbb{E}(\theta|\theta \in Q), \\ \omega(\pi, Q) &\triangleq \frac{\pi}{\pi + (1-\pi)\Pr(z \in Q)}. \end{aligned} \tag{22}$$

Proof of Lemma A.1. We proceed in 2 steps: Step 1 derives the lender's inference of the borrower's quality following data-withholding, Step 2 finds the optimal credit allocation.

Step 1. Let $Q \triangleq \{z \in [0,1] : m(z) = \varnothing\}$ be the set of borrowers that withhold data. Consider $\lambda \geq 0$, and let

$$\omega(\pi, Q) \triangleq \frac{\pi}{\pi + (1-\pi)\Pr(z \in Q)},$$

be the posterior probability that data-withholding is non-strategic. The lender's posterior beliefs over θ given $m = \varnothing$ are:

$$f_\lambda(\theta|\varnothing) = \omega(\pi, Q) f(\theta) + \big(1 - \omega(\pi, Q)\big) f_\lambda(\theta|z \in Q)$$

where $f(\theta) = 1$ is the prior distribution of θ, while

$$\begin{aligned} f_\lambda(\theta|z \in Q) &= \frac{\Pr\big(z \in Q|\theta\big) f(\theta)}{\Pr\big(z \in Q\big)}, \\ &= \frac{\displaystyle\int_Q \lambda\delta(\theta - z) + (1-\lambda)\ \mathrm{d}z}{\Pr\big(z \in Q\big)}, \\ &= \lambda \frac{\mathbb{1}\{\theta \in Q\}}{\Pr(\theta \in Q)} + (1-\lambda) \end{aligned}$$

is the posterior distribution given that data withholding is strategic ($\delta(\theta - z)$ is the Dirac's delta function defined in Equation (18) and satisfying property in Equation (19)). The expected θ given data withholding is

$$\begin{aligned} \mathbb{E}_\lambda(\theta|\varnothing) = &\int_0^1 \theta f_\lambda(\theta|\varnothing)\mathrm{d}\theta, \\ = &\,\omega(\pi, Q)\int_0^1 \theta\,\mathrm{d}\theta + \big(1 - \omega(\pi, Q)\big)\left[\lambda \int_Q \theta \frac{1}{\Pr(\theta \in Q)}\,\mathrm{d}\theta + (1-\lambda)\int_0^1 \theta\,\mathrm{d}\theta\right], \\ = &\,\omega(\pi, Q)\mathbb{E}(\theta) + \big(1 - \omega(\pi, Q)\big)\Big[\lambda\,\mathbb{E}(\theta|\theta \in Q) + (1-\lambda)\,\mathbb{E}(\theta)\Big], \\ = &\,\lambda\Big[\omega(\pi, Q)\tfrac{1}{2} + \big(1 - \omega(\pi, Q)\big)\mathbb{E}(\theta|\theta \in Q)\Big] + (1-\lambda)\tfrac{1}{2} \triangleq \mu_\lambda(\varnothing). \end{aligned}$$

Consider now $\lambda < 0$. For some $Q \subseteq [0,1]$ and some $y \in Q$, let $Q_r \triangleq \{1 - y | y \in Q\}$ be the reflection of Q over the axis $y = \frac{1}{2}$. The lender's posterior beliefs over θ given $m = \varnothing$ are:

$$f_\lambda(\theta|\varnothing) = \omega(\pi, Q) f(\theta) + \big(1 - \omega(\pi, Q)\big) f_\lambda(\theta|z \in Q)$$

where

$$\begin{aligned} f_\lambda(\theta|z \in Q) &= |\lambda| \frac{\mathbb{1}\{1 - \theta \in Q\}}{\Pr(z \in Q)} + (1 - |\lambda|), \\ &= |\lambda| \frac{\mathbb{1}\{\theta \in Q_r\}}{\Pr(\theta \in Q_r)} + (1 - |\lambda|), \end{aligned}$$

so that the expected θ given data withholding is

$$\begin{aligned}\mathbb{E}(\theta|\varnothing) = & \ |\lambda|\Big[\omega(\pi,Q)\tfrac{1}{2} + \big(1-\omega(\pi,Q)\big)\mathbb{E}(\theta|\theta\in Q_r)\Big] + (1-|\lambda|)\tfrac{1}{2}, \\ = & -\lambda\left[\omega(\pi,Q)\tfrac{1}{2} + \big(1-\omega(\pi,Q)\big)\big(1-\mathbb{E}(\theta|\theta\in Q)\big)\right] + (1+\lambda)\,\tfrac{1}{2}, \\ = & \ \lambda\left[\omega(\pi,Q)\tfrac{1}{2} + \big(1-\omega(\pi,Q)\big)\mathbb{E}(\theta|\theta\in Q)\right] + (1-\lambda)\,\tfrac{1}{2} = \mu_\lambda(\varnothing).\end{aligned}$$

Step 2. Suppose that after data withholding the lender provides credit, the interest rate solves:

$$\max_{x\in\mathbb{R}} \quad \Big[\mu_\lambda(\varnothing)\,x - 1\Big]^{\phi}\Big[\mu_\lambda(\varnothing)\big(X-x\big)\Big]^{1-\phi}.$$

Proceeding as in Lemma 1 we get the optimal interest rate:

$$x_\lambda(\varnothing) \triangleq \frac{1}{\mu_\lambda(\varnothing)} + \phi\left(X - \frac{1}{\mu_\lambda(\varnothing)}\right).$$

The lender provides credit only if her expected profits are strictly positive, i.e. only if:

$$\ell_\lambda(\varnothing) = \mathbb{1}\left\{\mu_\lambda(\varnothing) > \frac{1}{X}\right\}.$$

For future reference, I compute the lender's and borrower's expected profits and expected social welfare at the optimum, conditioning on $m=\varnothing$,

$$\begin{aligned}V_\lambda^L(\varnothing) &= \ell_\lambda(\varnothing)\Big[\phi\Big(\mu_\lambda(\varnothing)X-1\Big)\Big], \\ V_\lambda^B(\varnothing) &= \ell_\lambda(\varnothing)\Big[(1-\phi)\Big(\mu_\lambda(\varnothing)X-1\Big)+b\Big], \\ W_\lambda(\varnothing) &= \ell_\lambda(\varnothing)\Big[\mu_\lambda(\varnothing)X-1\Big],\end{aligned} \tag{23}$$

as well as the borrower's expected profits conditional on z

$$V_\lambda^B(\varnothing,z) = \ell_\lambda(\varnothing)\left[(1-\phi)\mu_\lambda(z)\left(X-\frac{1}{\mu_\lambda(\varnothing)}\right)+b\right]. \tag{24}$$

■

Proof of Proposition 1. Let $Q \triangleq \{z\in[0,1] : m(z)=\varnothing\}$ be the set of borrowers that withhold data, the borrower's utility from data sharing and withholding are respectively $V_\lambda^B(z)$ and $V_\lambda^B(\varnothing,z)$ from Equation (20) and Equation (24).

Consider first the case where $\lambda=0$ and the allocation rule is independent of data. No matter the data sharing strategy of the borrower there is credit provision at a flat interest rate, so that borrowers are indifferent between data sharing and withholding. Assuming that indifferent borrowers withholds data, the set of borrowers that withhold data is $Q=[0,1]$.

Consider now $\lambda > 0$ and suppose that the set of borrowers that withhold data is $Q = [0, q]$ with $q \in (0, 1)$. From Lemma A.1 we can compute the lender's estimate of the borrower's quality after data-withholding, $\mu_\lambda(z(\pi, q))$, where

$$z(\pi, q) = \frac{\pi}{\pi + (1-\pi)q}\frac{1}{2} + \left(1 - \frac{\pi}{\pi + (1-\pi)q}\right)\frac{q}{2},$$

as well as the optimal credit allocation. We want to find the borrower q that is indifferent between sharing and withholding data. Suppose that q is such that the lender provides credit both after data sharing and data withholding. If $\lambda \in (0, \overline{\lambda})$ with $\overline{\lambda} = 1 - 2/X$, this is the case for every z and $z(\pi, q)$ (see Lemma 1 and the discussion that follows). If $\lambda \in [\overline{\lambda}, 1]$, this requires that $q > r(\lambda)$ and $z(\pi, q) > r(\lambda)$ (see Lemma 1 and Lemma A.1) where $r(\lambda)$ is defined in Equation (8). Suppose that these conditions are satisfied, the indifferent borrower has to get the same surplus, i.e.

$$(1-\phi)\Big(\mu_\lambda(q)X - 1\Big) + b = (1-\phi)\mu_\lambda(q)\left(X - \frac{1}{\mu_\lambda(\varnothing)}\right) + b,$$

which simplifies to

$$q = z(\pi, q), \tag{25}$$

and whose unique positive solution is

$$q = \frac{\sqrt{\pi}(1-\sqrt{\pi})}{1-\pi} \triangleq \gamma(\pi) \in \left(0, \tfrac{1}{2}\right). \tag{26}$$

Hence, when $\gamma(\pi) > r(\lambda)$ the set of borrowers that withhold data is $Q = [0, \gamma(\pi)]$. When $\gamma(\pi) \leq r(\lambda)$, withholding data leads to credit rationing, so that all the borrowers with $z \in [0, r(\lambda)]$ are indifferent between data sharing and withholding. Assuming that indifferent borrowers withhold data, the set of borrowers that withhold data is $Q = [0, r(\lambda)]$.

When $\lambda \in [-1, 0)$, by a symmetric argument we have that when $1 - \gamma(\pi) < r(\lambda)$ the set of borrowers that withhold is $Q = [1 - \gamma(\pi), 1]$, while when $1 - \gamma(\pi) \geq r(\lambda)$ it is $Q = [r(\lambda), 1]$.

In summary, the equilibrium is described by the following data withholding set:

$$Q = \mathcal{G}(\lambda, \pi) \triangleq \begin{cases} \Big[0, \max\big\{r(\lambda); \gamma(\pi)\big\}\Big] & \text{if } \lambda \in (0, 1], \\ \Big[0, 1\Big] & \text{if } \lambda = 0, \\ \Big[\min\big\{r(\lambda); 1 - \gamma(\pi)\big\}, 1\Big] & \text{if } \lambda \in [-1, 0), \end{cases}$$

where $\gamma(\pi)$ and $r(\lambda)$ are given respectively by Equation (26) and Equation (8). ∎

Proof of Proposition 2. We proceed in 2 steps. Step 1 builds the borrower's expected utilities from data sharing and data withholding. Step 2 finds the equilibrium disclosure strategy.

Step 1. The λ-lender's allocation rule is described by Lemma 1, so the borrower's expected utility form data sharing is

$$V^B(z) = \int_{-1}^{1} V^B_\lambda(z)\,\frac{1}{2}\,\mathrm{d}\lambda$$

where $V^B_\lambda(z)$ is defined in Equation (20). Let

$$\widehat{\lambda}(z) \triangleq r^{-1}(z) = \frac{\frac{1}{2} - \frac{1}{X}}{\frac{1}{2} - z}, \tag{27}$$

be the inverse function of the rationing threshold defined in Equation (8). From Equation (7), we can distinguish 3 sets of borrowers: i) borrowers with $z \in [0, 1/X]$ are credit rationed whenever $\lambda \geq \widehat{\lambda}(z)$; ii) borrowers with $z \in (1/X, 1-1/X)$ get credit for every $\lambda \in [-1, 1]$; and iii) borrowers with $z \in [1-1/X, 1]$ are credit rationed whenever $\lambda \leq \widehat{\lambda}(z)$. Thus, the borrower's expected surplus from data sharing can be rewritten as

$$V^B(z) = \begin{cases} \displaystyle\int_{-1}^{\widehat{\lambda}(z)} \Big[(1-\phi)\Big(\mu_\lambda(z)X - 1\Big) + b\Big]\,\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \left[0, \dfrac{1}{X}\right] \\ \displaystyle\int_{-1}^{1} \Big[(1-\phi)\Big(\mu_\lambda(z)X - 1\Big) + b\Big]\,\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \left(\dfrac{1}{X}, 1 - \dfrac{1}{X}\right) \\ \displaystyle\int_{\widehat{\lambda}(z)}^{1} \Big[(1-\phi)\Big(\mu_\lambda(z)X - 1\Big) + b\Big]\,\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \left[1 - \dfrac{1}{X}, 1\right], \end{cases} \tag{28}$$

or, equivalently, as

$$V^B(z) = \begin{cases} p(z)\Big[(1-\phi)\Big(\mu_-(z)X - 1\Big) + b\Big] & \text{if } z \in \left[0, \dfrac{1}{X}\right] \\ (1-\phi)\left(\frac{1}{2}X - 1\right) + b & \text{if } z \in \left(\dfrac{1}{X}, 1 - \dfrac{1}{X}\right) \\ \Big(1 - p(z)\Big)\Big[(1-\phi)\Big(\mu_+(z)X - 1\Big) + b\Big] & \text{if } z \in \left[1 - \dfrac{1}{X}, 1\right], \end{cases} \tag{29}$$

where

$$\begin{aligned} p(z) &\triangleq \Pr\Big(\lambda \leq \widehat{\lambda}(z)\Big) = \frac{\widehat{\lambda}(z) + 1}{2}, \\ \mu_-(z) &\triangleq \mathbb{E}\Big(\mu_\lambda(z)\Big|\lambda \leq \widehat{\lambda}(z)\Big) = \frac{-1 + \widehat{\lambda}(z)}{2}\,z + \left(1 - \frac{-1 + \widehat{\lambda}(z)}{2}\right)\frac{1}{2}, \\ \mu_+(z) &\triangleq \mathbb{E}\Big(\mu_\lambda(z)\Big|\lambda \geq \widehat{\lambda}(z)\Big) = \frac{\widehat{\lambda}(z) + 1}{2}\,z + \left(1 - \frac{\widehat{\lambda}(z) + 1}{2}\right)\frac{1}{2}. \end{aligned}$$

Let $Q \triangleq \{z \in [0, 1] : m(z) = \varnothing\}$ be the set of borrowers that withhold data, the λ-lender's estimate of the borrower's quality $\mu_\lambda(\varnothing)$ and the optimal allocation rule $\big(\ell_\lambda(\varnothing), x_\lambda(\varnothing)\big)$ is given by Lemma A.1. The borrower's expected utility form data withholding is

$$V^B(\varnothing, z) = \int_{-1}^{1} V_\lambda^B(\varnothing, z)\,\frac{1}{2}\,\mathrm{d}\lambda,$$

where $V_\lambda^B(\varnothing, z)$ is defined in Equation (24). Proceeding as above we get the borrower's expected surplus from data withholding

$$V^B(\varnothing, z) = \begin{cases} \displaystyle\int_{-1}^{\widehat{\lambda}(z(\pi,Q))} \left[(1-\phi)\mu_\lambda(z)\left(X - \frac{1}{\mu_\lambda(\varnothing)}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z(\pi,Q) \in \left[0, \frac{1}{X}\right] \\ \displaystyle\int_{-1}^{1} \left[(1-\phi)\mu_\lambda(z)\left(X - \frac{1}{\mu_\lambda(\varnothing)}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z(\pi,Q) \in \left(\frac{1}{X}, 1-\frac{1}{X}\right) \\ \displaystyle\int_{\widehat{\lambda}(z(\pi,Q))}^{1} \left[(1-\phi)\mu_\lambda(z)\left(X - \frac{1}{\mu_\lambda(\varnothing)}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z(\pi,Q) \in \left[1-\frac{1}{X}, 1\right]. \end{cases}$$

Step 2. First, consider the case $\phi \geq 1 - 2b$. Suppose $Q = [0, 1]$, from Lemma A.1 (Equation (22)) we can compute the λ-lender's estimate of the borrower's quality following data withholding to get

$$\mu_\lambda(\varnothing) = z(\pi, Q) = \frac{1}{2} \in \left(\frac{1}{X}, 1 - \frac{1}{X}\right) \tag{30}$$

for all $\lambda \in [-1, 1]$. It follows that the borrower's utility from data withholding simplifies to (suppressing the dependency of z)

$$V^B(\varnothing) = (1-\phi)(\tfrac{1}{2}X - 1) + b \tag{31}$$

for all $z \in [0, 1]$.

We want to show that $V^B(\varnothing) \geq V^B(z)$ for all $z \in [0, 1]$. Note that $V^B(z)$ is continuous, symmetric around $z = 1/2$, constant and equal to $V^B(\varnothing)$ for $z \in (1/X, 1 - 1/X)$, and convex for $z \in [0, 1/X] \cup [1 - 1/X, 1]$. To show the convexity of $V^B(z)$ over $z \in [0, 1/X]$ note that its first derivative in this region is

$$\begin{aligned} \frac{\partial V^B(z)}{\partial z} &= \widehat{\lambda}'(z)\Big[(1-\phi)\Big(\mu_{\widehat{\lambda}(z)}(z)X - 1\Big) + b\Big]\frac{1}{2} + (1-\phi)X\frac{1}{2}\int_{-1}^{\widehat{\lambda}(z)} \lambda\,\mathrm{d}\lambda, \\ &= \widehat{\lambda}'(z)b\frac{1}{2} + (1-\phi)X\frac{1}{2}\int_{-1}^{\widehat{\lambda}(z)} \lambda\,\mathrm{d}\lambda, \end{aligned}$$

where the first line uses Leibnitz integral rule, and the second line follows from the fact that

$$\mu_{\widehat{\lambda}(z)}(z) = \frac{1}{2} + \frac{\frac{1}{2} - \frac{1}{X}}{\frac{1}{2} - z}\left(z - \frac{1}{2}\right) = \frac{1}{X};$$

while the second derivative is

$$\frac{\partial^2 V^B(z)}{\partial z^2} = \widehat{\lambda}''(z) b \frac{1}{2} + (1-\phi) X \frac{1}{2} \widehat{\lambda}'(z) \widehat{\lambda}(z) > 0,$$

since

$$\begin{aligned}
\widehat{\lambda}(z) &> 0, \\
\widehat{\lambda}'(z) &= \frac{\frac{1}{2} - \frac{1}{X}}{\left(\frac{1}{2} - z\right)^2} = \frac{1}{\frac{1}{2} - z} \widehat{\lambda}(z) > 0, \\
\widehat{\lambda}''(z) &= 2 \frac{\frac{1}{2} - \frac{1}{X}}{\left(\frac{1}{2} - z\right)^3} = 2 \frac{1}{\frac{1}{2} - z} \widehat{\lambda}'(z) > 0,
\end{aligned}$$

because every $z \in [0, \frac{1}{X}]$ is strictly below $\frac{1}{2}$ since by assumption 1 we have $\frac{1}{X} < \frac{1}{2}$. By symmetry, $V^B(z)$ is also convex over $[1 - \frac{1}{X}, 1]$.

It follows that $Q = [0, 1]$ is the equilibrium data-withholding set if borrowers with extreme $z \in 0, 1$ weakly prefers to withhold data. By symmetry of $V^B(z)$ it is sufficient that

$$V^B(\varnothing) = (1-\phi)\left(\frac{1}{2}X - 1\right) + b \geq \left(1 - \frac{1}{X}\right)\left[(1-\phi)\left(\left(1 + \frac{1}{X}\right)\frac{1}{2}X - 1\right) + b\right] = V^B(0).$$

This inequality simplifies to $\phi \geq 1 - 2b$, the parameters space we are considering.

Note also that any other putative equilibrium data withholding set of the form $Q = [0, q] \cup [1 - q, 1]$ with $q \in (0, 1/2)$ would lead to the estimated borrower's quality in Equation (30) and yield the profits in Equation (31) to a borrower withholding data, so that every borrower with $z \in (q, 1 - q)$ would deviate to data withholding.

Consider now $\phi < 1 - 2b$ and suppose $Q = [q, 1 - q]$ with $q \in (0, 1/X)$. From Lemma A.1 (Equation (22)) the λ-lender's estimate of the borrower's quality and the borrower's profits following data withholding are still given by Equation (30) and Equation (31), but for $\phi < 1 - 2b$ the borrower with $z = 0$ (and $z = 1$) strictly prefers to share data, as Equation (9) is not satisfied. The indifferent borrower q has to obtain the same expected surplus from both strategies, i.e.

$$V^B(\varnothing) = (1-\phi)\left(\frac{1}{2}X - 1\right) + b = p(q)\Big[(1-\phi)\Big(\mu_-(q)X - 1\Big) + b\Big] = V^B(q),$$

whose unique solution lower than $1/X$ is

$$q = \frac{1}{X}\left(1 - \frac{2b}{(1-\phi)}\right) \in \left(0, \tfrac{1}{X}\right). \tag{32}$$

By the symmetry of $V^B(z)$, we also have $V^B(\varnothing) = V^B(1 - q)$.

Since $q > 0$ if and only if $\phi < 1 - 2b$, the equilibrium can be compactly described by the data withholding set:

$$Q = \mathcal{H}(b, \phi) \triangleq [\eta(b, \phi), 1 - \eta(b, \phi)],$$

where

$$\eta(b,\phi) \triangleq \max\left\{0, \frac{1}{X}\left(1-\frac{2b}{1-\phi}\right)\right\}. \tag{33}$$

■

Proof of Proposition 3. We proceed in 4 steps: Step 1 and Step 2 derive the lender's expected profits when the lender employs a transparent and an opaque algorithm, respectively; Step 3 studies the graph of the lender's profits as a function of π; Step 4 compares the graphs.

Preliminaries. Consider a data sharing strategy of the form

$$m_Q(z) = \begin{cases} \varnothing & \text{if } z \in Q \\ z & \text{if } z \notin Q, \end{cases}$$

where $Q \in \{\mathcal{G}(\lambda,\pi), \mathcal{H}(b,\phi)\}$ is the data-withholding set induced by a specific transparency regime, and where $\mathcal{G}(\lambda,\pi)$ and $\mathcal{H}(b,\phi)$ are defined in Proposition 1 and Proposition 2, respectively. The lender's equilibrium profits from a transparency regime $\tau \in \{T, O\}$ inducing a data-withholding set Q are:

$$\begin{aligned} V_\tau^L(\pi) &\triangleq \frac{1}{2}\int_{-1}^{1} V_Q^L(\pi,\lambda)\,\mathrm{d}\lambda, \\ &= \int_0^1 V_Q^L(\pi,\lambda)\,\mathrm{d}\lambda, \end{aligned} \tag{34}$$

where

$$V_Q^L(\pi,\lambda) \triangleq \pi V_\lambda^L(\varnothing) + (1-\pi)\int_0^1 V_\lambda^L\big(m_Q(z)\big)\mathrm{d}z, \tag{35}$$

$V_\lambda^L(z)$ and $V_\lambda^L(\varnothing)$ are defined in Equation (20) and Equation (23), while the second line of Equation (34) follows from the lender's profit function being even in λ.

Step 1. Consider the lender's profits with a transparent algorithm.

If $\lambda = 0$ data is not used by the allocation rule and borrowers withhold data, i.e. $\mathcal{G}(\lambda,\pi) = [0,1]$, thus the lender's expected profits are equal to

$$V_{\mathcal{G}}^L(\pi,0) = \phi\left(\tfrac{1}{2}X - 1\right).$$

Now consider $\lambda \in \big(0, \widehat{\lambda}\big(\gamma(\pi)\big)\big)$ where $\widehat{\lambda}(\cdot)$ is defined in Equation (27). For these parameter values we either have that the allocation rule does not use data to ration credit (when $\lambda \in (0, \overline{\lambda})$), or the allocation rule rations credit but borrowers escape rationing by withholding data since $r(\lambda) < \gamma(\pi)$ (i.e. $\lambda \in \big[\overline{\lambda}, \widehat{\lambda}(\gamma(\pi))\big)$).

In both cases $\mathcal{G}(\lambda,\pi)=[0,\gamma(\pi)]$ (see Proposition 1). The lender's profits are:

$$\begin{aligned}
V_{\mathcal{G}}^{L}(\pi,\lambda) &= \Big(\pi+(1-\pi)\gamma(\pi)\Big)\phi\Big(\mu_\lambda(\varnothing)X-1\Big)+(1-\pi)\int_{\gamma(\pi)}^{1}\phi\Big(\mu_\lambda(z)X-1\Big)\mathrm{d}z,\\
&= \Big(\pi+(1-\pi)\gamma(\pi)\Big)\phi\Big(\mu_\lambda(\varnothing)X-1\Big)+\\
&+(1-\pi)(1-\gamma(\pi))\phi\left(\mu_\lambda\left(\tfrac{\gamma(\pi)+1}{2}\right)X-1\right),\\
&=\phi\left(\tfrac{1}{2}X-1\right),
\end{aligned}$$

where the last line follows from the fact that, in equilibrium, $\mu_\lambda(\varnothing)=\lambda\gamma(\pi)+(1-\lambda)\frac{1}{2}$ (see Proposition 1, Equation (25) and Equation (26)), and

$$\Big(\pi+(1-\pi)\gamma(\pi)\Big)\gamma(\pi)+(1-\pi)(1-\gamma(\pi))\frac{\gamma(\pi)+1}{2}=\frac{1}{2}.$$

Lastly, consider $\lambda\in\left[\widehat{\lambda}(\gamma(\pi)),1\right]$. We have $r(\lambda)\geq\gamma(\pi)$ and $\mathcal{G}(\lambda,\pi)=[r(\lambda),1]$, since borrowers that withhold data are credit rationed $\ell_\lambda(\varnothing)=0$. The lender's profits are:

$$V_{\mathcal{G}}^{L}(\pi,\lambda)=(1-\pi)\int_{r(\lambda)}^{1}\phi\Big(\mu_\lambda(z)X-1\Big)\mathrm{d}z.$$

Note that $\widehat{\lambda}\big(\gamma(\pi)\big)<1$ if and only if $\gamma(\pi)<1/X$, i.e. if $\pi<\left(\frac{\frac{1}{X}}{1-\frac{1}{X}}\right)^2\in(0,1)$, hence letting

$$\widetilde{\lambda}(\pi)\triangleq\min\left\{\widehat{\lambda}\big(\gamma(\pi)\big),1\right\},$$

and integrating $V_{\mathcal{G}}(\pi,\lambda)$ over λ we get the ex-ante expected profits from a transparent algorithm:

$$V_T^L(\pi)=\widetilde{\lambda}(\pi)\,\phi\left(\tfrac{1}{2}X-1\right)+(1-\pi)\int_{\widetilde{\lambda}(\pi)}^{1}\int_{r(\lambda)}^{1}\phi\Big(\mu_\lambda(z)X-1\Big)\,\mathrm{d}z\,\mathrm{d}\lambda. \tag{36}$$

Step 2. Consider the lender's profits when with an opaque algorithm when $\phi<1-2b$, i.e. $\eta(b,\phi)>0$.

If $\lambda\in[0,\overline{\lambda})$, where $\overline{\lambda}=1-2/X$, data is not used to ration credit, hence the lender's profits conditional on λ are:

$$\begin{aligned}
V_{\mathcal{H}}^{L}(\pi,\lambda) &= \Big(\pi+(1-\pi)\big(1-2\eta(b,\phi)\big)\Big)\phi\Big(\mu_\lambda(\varnothing)X-1\Big)+\\
&+(1-\pi)\left[\int_0^{\eta(b,\phi)}\phi\Big(\mu_\lambda(z)X-1\Big)\mathrm{d}z+\int_{1-\eta(b,\phi)}^{1}\phi\Big(\mu_\lambda(z)X-1\Big)\mathrm{d}z\right],\\
&= \Big(\pi+(1-\pi)\big(1-2\eta(b,\phi)\big)\Big)\phi\Big(\mu_\lambda(\varnothing)X-1\Big)+\\
&+(1-\pi)\eta(b,\phi)\phi\left[\mu_\lambda\left(\tfrac{\eta(b,\phi)}{2}\right)X-1+\mu_\lambda\left(\tfrac{2-\eta(b,\phi)}{2}\right)X-1\right],\\
&=\phi\left(\tfrac{1}{2}X-1\right),
\end{aligned}$$

where the last line follows from the fact that in equilibrium $\mu_\lambda(\varnothing) = \frac{1}{2}$ (see Proposition 2, Equation (30)) and

$$\mu_\lambda\left(\tfrac{\eta(b,\phi)}{2}\right) X - 1 + \mu_\lambda\left(\tfrac{2-\eta(b,\phi)}{2}\right) X - 1 = 2\left(\tfrac{1}{2}X - 1\right).$$

Consider next $\lambda \in [\overline{\lambda}, \widehat{\lambda}(\eta(b,\phi))]$ where $\widehat{\lambda}(\cdot)$ is defined in Equation (27). For these parameters we have $r(\lambda) \leq \eta(b,\phi)$, so that some of the borrowers sharing data are credit rationed. The lender's profits conditional on λ are:

$$\begin{aligned}
V^L_{\mathcal{H}}(\pi,\lambda) &= \pi\phi\Big(\mu_\lambda(\varnothing)X - 1\Big) + \\
&+ (1-\pi)\int_0^{\eta(b,\phi)} \mathbb{1}\{z > r(\lambda)\}\phi\Big(\mu_\lambda(z)X - 1\Big)\mathrm{d}z + \\
&+ (1-\pi)\Big[\int_{\eta(b,\phi)}^{1-\eta(b,\phi)} \phi\Big(\mu_\lambda(\varnothing)X - 1\Big)\mathrm{d}z + \int_{1-\eta(b,\phi)}^{1} \phi\Big(\mu_\lambda(z)X - 1\Big)\mathrm{d}z\Big], \\
&= \pi\phi\Big(\tfrac{1}{2}X - 1\Big) + (1-\pi)\int_{r(\lambda)}^{1} \phi\Big(\mu_\lambda(z)X - 1\Big)\mathrm{d}z,
\end{aligned}$$

where the fourth line follows from the fact that:

$$\begin{aligned}
\int_{\eta(b,\phi)}^{1-\eta(b,\phi)} \phi\Big(\mu_\lambda(\varnothing)X - 1\Big)\mathrm{d}z &= \int_{\eta(b,\phi)}^{1-\eta(b,\phi)} \phi\left(\tfrac{1}{2}X - 1\right)\mathrm{d}z \\
&= \int_{\eta(b,\phi)}^{1-\eta(b,\phi)} \phi\Big(\mu_\lambda(z)X - 1\Big)\mathrm{d}z.
\end{aligned} \tag{37}$$

Lastly, consider $\lambda \in (\widehat{\lambda}(\eta(b,\phi)), 1]$, that is parameters such that $r(\lambda) > \eta(b,\phi)$ and all the borrowers that share data $z < \eta(b,\phi)$ are credit rationed. The lender's profits are:

$$\begin{aligned}
V^L_{\mathcal{H}}(\pi,\lambda) &= \pi\phi\Big(\mu_\lambda(\varnothing)X - 1\Big) + \\
&+ (1-\pi)\Big[\int_{\eta(b,\phi)}^{1-\eta(b,\phi)} \phi\Big(\mu_\lambda(\varnothing)X - 1\Big)\mathrm{d}z + \int_{1-\eta(b,\phi)}^{1} \phi\Big(\mu_\lambda(z)X - 1\Big)\mathrm{d}z\Big], \\
&= \pi\phi\left(\tfrac{1}{2}X - 1\right) + (1-\pi)\int_{\eta(b,\phi)}^{1} \phi\Big(\mu_\lambda(z)X - 1\Big)\mathrm{d}z,
\end{aligned}$$

where the third line follows from Equation (37).

Integrating $V^L_{\mathcal{H}}(b,\phi)$ over λ, we get the lender's ex-ante expected profits from an opaque algorithm are:

$$\begin{aligned}
V^L_O(\pi) &= \overline{\lambda}\,\phi\left(\tfrac{1}{2}X - 1\right) + \\
&+ \int_{\overline{\lambda}}^{\widehat{\lambda}(\eta(b,\phi))} \pi\phi\left(\tfrac{1}{2}X - 1\right) + (1-\pi)\int_{r(\lambda)}^{1} \phi\Big(\mu_\lambda(z)X - 1\Big)\,\mathrm{d}z\,\mathrm{d}\lambda + \\
&+ \int_{\widehat{\lambda}(\eta(b,\phi))}^{1} \pi\phi\left(\tfrac{1}{2}X - 1\right) + (1-\pi)\int_{\eta(b,\phi)}^{1} \phi\Big(\mu_\lambda(z)X - 1\Big)\,\mathrm{d}z\,\mathrm{d}\lambda.
\end{aligned} \tag{38}$$

Note that when $\phi \geq 1-2b$ we have $\eta(b,\phi)=0$ and $\widehat{\lambda}(\eta(b,\phi))=\overline{\lambda}$, hence the lender's profits from Equation (38) simplifies to $V_O^L(\pi)=\phi\left(\frac{1}{2}X-1\right)$.

Step 3. We now study the graph of the lender's profits in Equation (36) and Equation (38). For an arbitrary function $y:(0,1)\to\mathbb{R}$, we will use the simplified notation $y(0)=\lim_{\pi\to 0} y(\pi)$ and $y(1)=\lim_{\pi\to 1} y(\pi)$ to denote the limiting values of $y(\pi)$ as π approaches 0 and 1, respectively.

First, since $\widetilde{\lambda}(1)=1$ and $\widetilde{\lambda}(0)=\overline{\lambda}$, we have that

$$V_O^L(1)=V_T^L(1)=\phi\left(\tfrac{1}{2}X-1\right),$$

and

$$V_T^L(0)-V_O^L(0)=-\int_{\widehat{\lambda}(\eta(b,\phi))}^{1}\int_{\eta(b,\phi)}^{r(\lambda)}\phi\Big(\mu_\lambda(z)X-1\Big)\,\mathrm{d}z\,\mathrm{d}\lambda>0,$$

where this uses the fact that

$$\left(\tfrac{1}{2}X-1\right)=\int_0^1\Big(\mu_\lambda(z)X-1\Big)\,\mathrm{d}z, \tag{39}$$

and the inequality follows from the fact that the integrand is negative for $z\leq r(\lambda)$.

Second, $V_O^L(\pi)$ is flat and equal to $\phi\left(\frac{1}{2}X-1\right)=V_L^O(1)$ if $\phi\geq 1-2b$ (i.e. when $\eta(b,\phi)=0$ and $\widehat{\lambda}(\eta(b,\phi))=\overline{\lambda}$) and is strictly decreasing if $\phi<1-2b$ (i.e. when $\eta(b,\phi)>0$ and $\widehat{\lambda}(\eta(b,\phi))>\overline{\lambda}$), since

$$\begin{aligned}
\frac{\partial V_O^L}{\partial\pi}&=\int_{\overline{\lambda}}^{\widehat{\lambda}(\eta(b,\phi))}\phi\left(\tfrac{1}{2}X-1\right)-\int_{r(\lambda)}^{1}\phi\Big(\mu_\lambda(z)X-1\Big)\,\mathrm{d}z\,\mathrm{d}\lambda+\\
&+\int_{\widehat{\lambda}(\eta(b,\phi))}^{1}\phi\left(\tfrac{1}{2}X-1\right)-\int_{\eta(b,\phi)}^{1}\phi\Big(\mu_\lambda(z)X-1\Big)\,\mathrm{d}z\,\mathrm{d}\lambda,\\
&=\int_{\overline{\lambda}}^{1}\int_0^{\min\{r(\lambda),\eta(b,\phi)\}}\phi\Big(\mu_\lambda(z)X-1\Big)\,\mathrm{d}z\,\mathrm{d}\lambda<0,
\end{aligned}$$

where the third line follows from Equation (39) and the fact that $r(\lambda)\geq\eta(b,\phi)$ if and only if $\lambda\geq\widehat{\lambda}(\eta(b,\phi))$, while the inequality follows from the integrand being negative for $z\leq r(\lambda)$.

Third, $V_T^L(\pi)$ is flat and equal to $\phi\left(\frac{1}{2}X-1\right)=V_T^L(1)$ if $\pi\geq\left(\frac{1}{X}/(1-\frac{1}{X})\right)^2\in(0,1)$ (i.e. when $\widetilde{\lambda}(\pi)=1$) and is strictly decreasing for $\pi<\left(\frac{1}{X}/(1-\frac{1}{X})\right)^2$ (i.e. when $\widetilde{\lambda}(\pi)=\widehat{\lambda}(\gamma(\pi))<1$). To prove that $V_T^L(\pi)$ is decreasing for $\pi<\left(\frac{1}{X}/(1-\frac{1}{X})\right)^2$, rewrite it as

$$V_T^L(\pi)=\widehat{\lambda}(\gamma(\pi))\,\phi\left(\tfrac{1}{2}X-1\right)+(1-\pi)\int_{\widehat{\lambda}(\gamma(\pi))}^{1}K(\lambda)\,\mathrm{d}\lambda$$

where

$$\begin{aligned}
K(\lambda)&\triangleq\int_{r(\lambda)}^{1}\phi\Big(\mu_\lambda(z)X-1\Big)\;\mathrm{d}z,\\
&=(1-r(\lambda))\phi\left(\mu_\lambda\left(\tfrac{r(\lambda)+1}{2}\right)X-1\right),
\end{aligned} \tag{40}$$

and $\mu_\lambda(\cdot)$ is defined in Equation (2). The first derivative of $V_T^L(\pi)$ is:

$$
\begin{aligned}
\frac{\partial V_T^L}{\partial \pi} &= \widehat{\lambda}'(\gamma(\pi))\gamma'(\pi)\left[\phi\left(\tfrac{1}{2}X-1\right)-(1-\pi)K\big(\widehat{\lambda}(\gamma(\pi))\big)\right]-\int_{\widehat{\lambda}(\gamma(\pi))}^{1}K(\lambda)\mathrm{d}\lambda, \\
&= -\int_{\widehat{\lambda}(\gamma(\pi))}^{1}K(\lambda)\mathrm{d}\lambda<0,
\end{aligned}
\tag{41}
$$

where the first line uses Leibnitz integral rule, the second line follows from the term in square brackets being nil, while the inequality follows from the fact that the integrand Equation (40) is strictly positive for $z>r(\lambda)$. To see that the term in square brackets in the first line of Equation (41) is nil, substitute $\widehat{\lambda}(\gamma(\pi))$ in Equation (40) and note that

$$
r\big(\widehat{\lambda}(\gamma(\pi))\big)=\gamma(\pi), \tag{42}
$$

since $\widehat{\lambda}(\cdot)=r^{-1}(\cdot)$ by definition (see Equation (27)), implying that

$$
\begin{aligned}
\mu_{\widehat{\lambda}(\gamma(\pi))}\left(\tfrac{r(\widehat{\lambda}(\gamma(\pi)))+1}{2}\right)X-1 &= \left(\widehat{\lambda}(\gamma(\pi))\frac{\gamma(\pi)+1}{2}+\big(1-\widehat{\lambda}(\gamma(\pi))\big)\frac{1}{2}\right)X-1, \\
&= \left(\tfrac{1}{2}X-1\right)+\frac{1}{2}\widehat{\lambda}(\gamma(\pi))\gamma(\pi)X, \\
&= \left(\tfrac{1}{2}X-1\right)+\frac{\sqrt{\pi}}{1-\sqrt{\pi}}\left(\tfrac{1}{2}X-1\right)>0;
\end{aligned}
\tag{43}
$$

hence the term in square brackets in Equation (41) simplifies to

$$
\phi\left[\left(\tfrac{1}{2}X-1\right)-(1-\pi)(1-\gamma(\pi))\left(\left(\tfrac{1}{2}X-1\right)+\frac{\sqrt{\pi}}{1-\sqrt{\pi}}\left(\tfrac{1}{2}X-1\right)\right)\right]=0,
$$

since $(1-\pi)(1-\gamma(\pi))=1-\sqrt{\pi}$.

Step 4. Let

$$
\widehat{\pi}(b,\phi)\triangleq\left\{\pi\in\left(0,\left(\frac{\frac{1}{X}}{1-\frac{1}{X}}\right)^2\right)\middle|\,V_T^L(\pi)=V_O^L(\pi)\ \text{ when}\,\phi<1-2b\right\} \tag{44}
$$

be the π that solves $V_T^L(\pi)=V_O^L(\pi)$ where $V_T^L(\pi)$ is given by Equation (36) with $\widetilde{\lambda}(\pi)=\widehat{\lambda}(\gamma(\pi))$, while $V_O^L(\pi)$ is given by Equation (38) with $\eta(b,\phi)>0$. From the shapes of $V_T^L(\pi)$ and $V_O^L(\pi)$ described in Step 3, it follows that:

- if $\phi\geq 1-2b$, we have
  - $V_T^L(\pi)>V_O^L(\pi)$ for $\pi<\left(\frac{1}{X}/(1-\frac{1}{X})\right)^2$,
  - $V_T^L(\pi)=V_O^L(\pi)$ for $\pi\geq\left(\frac{1}{X}/(1-\frac{1}{X})\right)^2$,
- if $\phi<1-2b$, we have

- $V_T^L(\pi) > V_O^L(\pi)$ for $\pi < \widehat{\pi}(b,\phi)$,
- $V_T^L(\pi) = V_O^L(\pi)$ for $\pi = \widehat{\pi}(b,\phi)$,
- $V_T^L(\pi) < V_O^L(\pi)$ for $\pi > \widehat{\pi}(b,\phi)$.

■

Proof of Lemma 2. Given a risk bucket r_i, the lender allocates credit as described in Lemma 1. In particular, expected credit quality is

$$\mathbb{E}(\theta|r_i) = \lambda\frac{z_{i-1}+z_i}{2} + (1-\lambda)\tfrac{1}{2} \triangleq \mu_\lambda\left(r_i\right),$$

while the allocation rule is:

$$x_\lambda(r_i) = \frac{1}{\mu_\lambda(r_i)} + \phi\left(X - \frac{1}{\mu_\lambda(r_i)}\right),$$
$$\ell_\lambda(r_i) = \mathbb{1}\left\{\mu_\lambda(r_i) > \frac{1}{X}\right\}.$$

Suppose $\lambda > 0$, the welfare-optimal data segmentation solves:

$$\max_{Z} \quad \sum_{i=1}^{n}\left(z_i - z_{i-1}\right)\ell_\lambda(r_i)\big[\mu_\lambda\left(r_i\right)X - 1 + b\big].$$

By the revelation principle, we can pool all the potential risk buckets r_i inducing the same credit provision decision in a unique signal and interpret signals as incentive-compatible action recommendations. The planner's problem simplifies to:

$$\begin{aligned} \max_{z_1\in[0,1]} \quad & \left(1-z_1\right)\left[\mu_\lambda\left(\tfrac{z_1+1}{2}\right)X - 1 + b\right] \\ \text{s.t.} \quad & \mu_\lambda\left(\tfrac{z_1+1}{2}\right) > \frac{1}{X} \\ & \mu_\lambda\left(\tfrac{0+z_1}{2}\right) \leq \frac{1}{X}. \end{aligned} \tag{45}$$

The objective function is concave and the first order condition is solved for

$$z_1 = \frac{1}{2} - \frac{1}{\lambda}\left(\frac{1}{2} - \frac{1}{X}\right) - \frac{b}{\lambda X} = r(\lambda) - \frac{b}{\lambda X},$$

which is positive as long as $b \leq 1 - \frac{1}{2}(1-\lambda)X$. One can easily show that the constraints are satisfied for

$$z_1 = \max\left\{r(\lambda) - \frac{b}{\lambda X}, 0\right\},$$

hence this is the solution of the problem in Equation (45) for $\lambda > 0$.

For $\lambda < 0$, relabeling action recommendations, we get

$$z_1 = \min \left\{ r(\lambda) - \frac{b}{\lambda X}, 1 \right\}.$$

When $\lambda = 0$, data segmentation is irrelevant for both the social planner and the lender. ■

Proof of Proposition 4. We proceed in 3 steps: Step 1 derives the social welfare when the lender employs a transparent and an opaque algorithm; Step 2 studies the graph of social welfare in the two regimes as a function of π; Step 3 compares the graphs.

Preliminaries. Consider a data sharing strategy of the form

$$m_Q(z) = \begin{cases} \varnothing & \text{if } z \in Q \\ z & \text{if } z \notin Q, \end{cases}$$

where $Q \in \{\mathcal{G}(\lambda, \pi), \mathcal{H}(b, \phi)\}$ is the data-withholding set induced by a specific transparency regime, and where $\mathcal{G}(\lambda, \pi)$ and $\mathcal{H}(b, \phi)$ are defined in Proposition 1 and Proposition 2, respectively. The equilibrium social welfare from a transparency regime $\tau \in \{T, O\}$ inducing a data-withholding set Q is:

$$\begin{aligned} W_\tau(\pi) &\triangleq \frac{1}{2} \int_{-1}^{1} W_Q(\pi, \lambda) \, \mathrm{d}\lambda, \\ &= \int_0^1 W_Q(\pi, \lambda) \, \mathrm{d}\lambda, \end{aligned} \tag{46}$$

where

$$W_Q(\pi, \lambda) \triangleq \pi W_\lambda(\varnothing) + (1 - \pi) \int_0^1 W_\lambda\big(m_Q(z)\big) \mathrm{d}z, \tag{47}$$

$W_\lambda(z)$ and $W_\lambda(\varnothing)$ are defined in Equation (20) and Equation (23), while the second line of Equation (46) follows from the social surplus function being even in λ.

Step 1. Proceeding as in Step 1 and Step 2 of Proposition 3, we get the social welfare from a transparent and an opaque algorithm. These are, respectively

$$W_T(\pi) = \widetilde{\lambda}(\pi) \left(\frac{1}{2} X - 1 + b \right) + (1 - \pi) \int_{\widetilde{\lambda}(\pi)}^{1} \int_{r(\lambda)}^{1} \Big(\mu_\lambda(z) X - 1 + b \Big) \, \mathrm{d}z \, \mathrm{d}\lambda, \tag{48}$$

and

$$\begin{aligned} W_O(\pi) = \overline{\lambda} \left(\frac{1}{2} X - 1 + b \right) + \\ + \int_{\overline{\lambda}}^{\widehat{\lambda}(\eta(b,\phi))} \left[\pi \left(\frac{1}{2} X - 1 + b \right) + (1 - \pi) \int_{r(\lambda)}^{1} \Big(\mu_\lambda(z) X - 1 + b \Big) \, \mathrm{d}z \right] \mathrm{d}\lambda + \\ + \int_{\widehat{\lambda}(\eta(b,\phi))}^{1} \left[\pi \left(\frac{1}{2} X - 1 + b \right) + (1 - \pi) \int_{\eta(b,\phi)}^{1} \Big(\mu_\lambda(z) X - 1 + b \Big) \, \mathrm{d}z \right] \mathrm{d}\lambda. \end{aligned} \tag{49}$$

Step 2. We now study the graph of the lender's profits in Equation (48) and Equation (49).

Step 2a. We first evaluate $W_T(\pi)$ and $W_O(\pi)$ when π approaches 0 and 1. For an arbitrary function $y : (0,1) \to \mathbb{R}$, we will use the simplified notation $y(0) = \lim_{\pi\to 0} y(\pi)$ and $y(1) = \lim_{\pi \to 1} y(\pi)$ to denote the limiting values of $y(\pi)$ as π approaches 0 and 1, respectively.

First, since $\widetilde{\lambda}(1) = 1$ and $\widetilde{\lambda}(0) = \overline{\lambda}$, we have that

$$W_O(1) = W_T(1) = \tfrac{1}{2}X - 1 + b,$$

and, using Equation (39)

$$W_O(0) - W_T(0) = \int_{\widehat{\lambda}\left(\eta(b,\phi)\right)}^{1} \int_{\eta(b,\phi)}^{r(\lambda)} \mu_\lambda(z)X - 1 + b \;\, \mathrm{d}z \, \mathrm{d}\lambda \triangleq Z(b,\phi). \tag{50}$$

The sign of $Z(b,\phi)$ depends on its arguments. Rewrite this function as

$$Z(b,\phi) = \int_{\widehat{\lambda}\left(\eta(b,\phi)\right)}^{1} Y(b,\phi,\lambda)\,\mathrm{d}\lambda$$

where

$$Y(b,\phi,\lambda) \triangleq \int_{\eta(b,\phi)}^{r(\lambda)} \mu_\lambda(z)X - 1 + b \;\, \mathrm{d}z. \tag{51}$$

Note that $Z(b,\phi)$ is continuous, constant in ϕ for $\phi \geq 1-2b$ (i.e. when $\eta(b,\phi)=0$ and $\widehat{\lambda}(\eta(b,\phi)) = \overline{\lambda}$) and equal to

$$Z_-(b) \triangleq \int_{\overline{\lambda}}^{1} \int_{0}^{r(\lambda)} \mu_\lambda(z)X - 1 + b \, \mathrm{d}z \, \mathrm{d}\lambda, \tag{52}$$

while it is strictly decreasing in ϕ for $\phi < 1-2b$, since

$$\begin{aligned}
\frac{\partial Z}{\partial \phi} &= -\frac{\partial \widehat{\lambda}}{\partial \eta}\frac{\partial \eta}{\partial \phi} Y\big(b,\phi,\widehat{\lambda}(\eta(b,\phi))\big) + \int_{\widehat{\lambda}\left(\eta(b,\phi)\right)}^{1} \frac{\partial Y}{\partial \phi}\,\mathrm{d}\lambda, \\
&= \int_{\widehat{\lambda}\left(\eta(b,\phi)\right)}^{1} \frac{\partial Y}{\partial \phi}\,\mathrm{d}\lambda, \\
&= \frac{2b}{X(1-\phi)^2} \int_{\widehat{\lambda}\left(\eta(b,\phi)\right)}^{1} \mu_\lambda\big(\eta(b,\phi)\big)X - 1 + b \, \mathrm{d}\lambda, \\
&= \frac{2b}{X(1-\phi)^2}\big(1-\widehat{\lambda}(\eta(b,\phi))\big)\left[\left(\tfrac{\widehat{\lambda}(\eta(b,\phi))+1}{2}\eta(b,\phi) + \left(1 - \tfrac{\widehat{\lambda}(\eta(b,\phi))+1}{2}\right)\frac{1}{2}\right)X - 1 + b\right], \\
&= \frac{2b}{X(1-\phi)^2}\big(1-\widehat{\lambda}(\eta(b,\phi))\big)\left[\left(\frac{1}{X} - \frac{b}{(1-\phi)X}\right)X - 1 + b\right] < 0,
\end{aligned} \tag{53}$$

where the first line uses the Leibnitz integral rule, the second lines follows from the fact that $Y\big(b,\phi,\widehat{\lambda}(\eta(b,\phi))\big) = 0$ since $r\big(\widehat{\lambda}\big(\eta(b,\phi)\big)\big) = \eta(b,\phi)$ because of Equation (42), while the inequality follows from the fact that the term in square brackets is strictly negative.

Moreover, $Z(b,\phi)$ is strictly increasing in b. This is obvious when $\phi \geq 1-2b$ and $Z(b,\phi) = Z_{-}(b)$, while for $\phi < 1-2b$ we have

$$\begin{aligned}
\frac{\partial Z}{\partial b} &= -\frac{\partial \widehat{\lambda}}{\partial \eta}\frac{\partial \eta}{\partial b} Y\big(b,\phi,\widehat{\lambda}(\eta(b,\phi))\big) + \int_{\widehat{\lambda}(\eta(b,\phi))}^{1} \left[-\frac{\partial \eta}{\partial b}\big(\mu_{\lambda}(\eta(b,\phi))X - 1 + b\big) + \int_{\eta(b,\phi)}^{r(\lambda)} \mathrm{d}z \right] \mathrm{d}\lambda, \\
&= \frac{2}{X(1-\phi)} \int_{\widehat{\lambda}(\eta(b,\phi))}^{1} \Big[\mu_{\lambda}(\eta(b,\phi))X - 1 + b\Big] \mathrm{d}\lambda + \int_{\widehat{\lambda}(\eta(b,\phi))}^{1} \int_{\eta(b,\phi)}^{r(\lambda)} \mathrm{d}z \, \mathrm{d}\lambda, \\
&= \frac{2}{X(1-\phi)} \left[\Big(1 - \widehat{\lambda}(\eta(b,\phi))\Big) \left(\tfrac{1}{2}X - 1 + b + \frac{\widehat{\lambda}(\eta(b,\phi)) + 1}{2} \eta(b,\phi) \right) \right] + \int_{\widehat{\lambda}(\eta(b,\phi))}^{1} \int_{\eta(b,\phi)}^{r(\lambda)} \mathrm{d}z \, \mathrm{d}\lambda > 0,
\end{aligned} \tag{54}$$

where the first lines uses Leibnitz integral rule, the second lines follows from $Y\big(\phi,\widehat{\lambda}\big(\eta(b,\phi)\big)\big) = 0$ since $r\big(\widehat{\lambda}\big(\eta(b,\phi)\big)\big) = \eta(b,\phi)$ because of Equation (42), while the inequality in the last line follows from the following facts: i) the term in square brackets is positive because $\widehat{\lambda}(\eta(b,\phi)) \in (0,1)$ and $\frac{1}{2}X - 1 > 0$ by Assumption 1; and ii) the double integral is positive since for $\lambda > \widehat{\lambda}(\eta(b,\phi))$ we have $r(\lambda) > \eta(b,\phi)$.

Moreover, note that $Z(0,\phi) = 0$ since $\widehat{\lambda}(\eta(0,\phi)) = 1$, and that $Z_{-}(b) > 0$ if and only if $b > b^{\circ}(X)$ where

$$b^{\circ}(X) \triangleq -\frac{\displaystyle\int_{\overline{\lambda}}^{1} \int_{0}^{r(\lambda)} \mu_{\lambda}(z)X - 1 \, \mathrm{d}z \, \mathrm{d}\lambda}{\displaystyle\int_{\overline{\lambda}}^{1} \int_{0}^{r(\lambda)} \mathrm{d}z \, \mathrm{d}\lambda} \in \left(\tfrac{1}{4}, \tfrac{1}{3}\right). \tag{55}$$

Let

$$\phi^{\circ}(b) \triangleq \left\{ \phi \in (0, 1-2b) \,\middle|\, Z(b,\phi) \;\; \text{when}\, \phi < 1-2b \right\} \tag{56}$$

be the ϕ that solves $Z(b,\phi) = 0$ where $Z(b,\phi)$ is given by Equation (50) with $\eta(b,\phi) > 0$.

It follows that:

- if $b > b^{\circ}(X)$, then $Z(b,\phi) > 0$ for every $\phi \in (0,1)$;
- if $b = b^{\circ}(X)$, then $Z(b,\phi) > 0$ for $\phi \in (0, 1-2b)$ and $Z(b,\phi) = 0$ for $\phi \in [1-2b, 1)$;
- if $b < b^{\circ}(X)$, then $Z(b,\phi) > 0$ if $\phi \in (0, \phi^{\circ}(b))$, $Z(b,\phi) = 0$ if $\phi = \phi^{\circ}(b)$, and $Z(b,\phi) < 0$ if $\phi \in (\phi^{\circ}(b), 1)$.

Step 2b. We now study the graph of $W_T(\pi)$ and $W_O(\pi)$ over the interior of the interval $(0,1)$.

First, note that $W_O(\pi)$ is flat and equal to $\frac{1}{2}X - 1 + b = W_O(1)$ if $\phi \geq 1-2b$ (i.e. when $\eta(b,\phi) = 0$

and $\widehat{\lambda}(\eta(b,\phi)) = \overline{\lambda})$ and is strictly decreasing if $\phi < 1-2b$ (i.e. when $\eta(b,\phi) > 0$ and $\widehat{\lambda}(\eta(b,\phi)) > \overline{\lambda})$ since

$$\begin{aligned}
\frac{\partial W_O}{\partial \pi} &= \int_{\overline{\lambda}}^{\widehat{\lambda}(\eta(b,\phi))} \left[\left(\tfrac{1}{2}X - 1 + b\right) - \int_{r(\lambda)}^{1} \mu_\lambda(z)X - 1 + b \; \mathrm{d}z \right] \mathrm{d}\lambda + \\
&+ \int_{\widehat{\lambda}(\eta(b,\phi))}^{1} \left[\left(\tfrac{1}{2}X - 1 + b\right) - \int_{\eta(b,\phi)}^{1} \mu_\lambda(z)X - 1 + b \; \mathrm{d}z \right] \mathrm{d}\lambda, \\
&= \int_{\overline{\lambda}}^{1} \int_{0}^{\min\{r(\lambda),\eta(b,\phi)\}} \left[\mu_\lambda(z)X - 1 + b\right] \mathrm{d}z \, \mathrm{d}\lambda, \\
&< \int_{\overline{\lambda}}^{1} \int_{0}^{\eta(b,\phi)} \left[\mu_\lambda(z)X - 1 + b\right] \mathrm{d}z \, \mathrm{d}\lambda, \\
&= (1-\overline{\lambda})\eta(b,\phi) \left[\left(\frac{\overline{\lambda}+1}{2} \frac{\eta(b,\phi)}{2} + \left(1 - \frac{\overline{\lambda}+1}{2}\right) \frac{1}{2} \right) X - 1 + b \right] < 0
\end{aligned} \tag{57}$$

where the third line follows from Equation (39) and the fact that $r(\lambda) \geq \eta(b,\phi)$ if and only if $\lambda \geq \widehat{\lambda}(\eta(b,\phi))$, while the last inequality follows from the fact that the term in square brackets is strictly negative. To see this note that the term in square brackets is linear in b; it is increasing in b for $\phi < 1/X$ and equal to $-1/2 + (1-\phi)/2 < 0$ for the highest value of b among the parameters considered, i.e. $b = (1-\phi)/2$; it is decreasing in b for $\phi > 1/X$ and equal to $-1/(2X)$ for the lowest value of b, i.e. $b = 0$; it is constant in b for $\phi = 1/X$ and equal to $-1/(2X) < 0$.

Second, note that $W_T(\pi)$ is flat and equal to $\frac{1}{2}X - 1 + b = W_T(1)$ if $\pi \geq \left(\frac{1}{X}/(1-\frac{1}{X})\right)^2 \in (0,1)$ (i.e. when $\widetilde{\lambda}(\pi) = 1$) and is strictly convex for $\pi < \left(\frac{1}{X}/(1-\frac{1}{X})\right)^2$ (i.e. when $\widetilde{\lambda}(\pi) = \widehat{\lambda}(\gamma(\pi)) < 1$) approaching the flat part from below. To prove that $W_T(\pi)$ is convex for $\pi < \left(\frac{1}{X}/(1-\frac{1}{X})\right)^2$ and approaches the flat part from below, rewrite it as

$$W_T(\pi) = \widehat{\lambda}(\gamma(\pi)) \; \left[\tfrac{1}{2}X - 1 + b\right] + (1-\pi) \int_{\widehat{\lambda}(\gamma(\pi))}^{1} H(\lambda) \, \mathrm{d}\lambda$$

where

$$\begin{aligned}
H(\lambda) &\triangleq \int_{r(\lambda)}^{1} \left[\mu_\lambda(z)X - 1 + b\right] \; \mathrm{d}z, \\
&= (1 - r(\lambda)) \left[\mu_\lambda\left(\tfrac{r(\lambda)+1}{2}\right) X - 1 + b \right].
\end{aligned} \tag{58}$$

The first derivative of $W_T(\pi)$ is :

$$\begin{aligned}
\frac{\partial W_T}{\partial \pi} &= \widehat{\lambda}'(\gamma(\pi))\gamma'(\pi) \left[\left(\tfrac{1}{2}X - 1 + b\right) - (1-\pi)\, H\big(\widehat{\lambda}(\gamma(\pi))\big) \right] - \int_{\widehat{\lambda}(\gamma(\pi))}^{1} H(\lambda) \, \mathrm{d}\lambda, \\
&= \widehat{\lambda}'(\gamma(\pi))\gamma'(\pi) \Big[1 - (1-\pi)(1-\gamma(\pi))\Big] b - \int_{\widehat{\lambda}(\gamma(\pi))}^{1} H(\lambda) \, \mathrm{d}\lambda, \\
&= \frac{\overline{\lambda}}{(1-\sqrt{\pi})^2} \, b - \int_{\widehat{\lambda}(\gamma(\pi))}^{1} H(\lambda) \, \mathrm{d}\lambda,
\end{aligned}$$

where the first line uses the Leibnitz integral rule, the second line follows from plugging $\widehat{\lambda}(\gamma(\pi))$ in Equation (58) and using Equation (42) and Equation (43) and the fact that $(1-\pi)(1-\gamma(\pi)) = 1-\sqrt{\pi}$, while the last line follows from the fact that

$$\begin{aligned} \widehat{\lambda}'(\gamma(\pi))\gamma'(\pi) &= \frac{\overline{\lambda}}{\sqrt{\pi}(1-\sqrt{\pi})^2} > 0, \\ (1-\pi)(1-\gamma(\pi)) &= 1-\sqrt{\pi}. \end{aligned} \tag{59}$$

Note that the left derivative of $W_T(\pi)$ at $\pi = \left(\frac{1}{X}/(1-\frac{1}{X})\right)^2$ is

$$\frac{\partial W_T}{\partial \pi}\Big|_{\pi=\left(\frac{\frac{1}{X}}{1-\frac{1}{X}}\right)^2} = \frac{\overline{\lambda}}{\left(1-\frac{\frac{1}{X}}{1-\frac{1}{X}}\right)^2} b > 0, \tag{60}$$

since $\widehat{\lambda}(\gamma(\pi)) = 1$ for $\pi = \left(\frac{1}{X}/(1-\frac{1}{X})\right)^2$.

The second derivative of $W_T(\pi)$ is

$$\begin{aligned} \frac{\partial^2 W_T}{\partial \pi^2} &= \frac{\overline{\lambda}}{\sqrt{\pi}(1-\sqrt{\pi})^3} b + \widehat{\lambda}'(\gamma(\pi))\gamma'(\pi)\, H\big(\widehat{\lambda}(\gamma(\pi))\big), \\ &= \frac{\overline{\lambda}}{\sqrt{\pi}(1-\sqrt{\pi})^2}\left[b\left(\frac{1}{1-\sqrt{\pi}} + 1 - \gamma(\pi)\right) + (1-\gamma(\pi))\left(\tfrac{1}{2}X - 1\right)\left(1 + \frac{\sqrt{\pi}}{1-\sqrt{\pi}}\right)\right] > 0, \end{aligned} \tag{61}$$

where the first line follows from the Leibnitz integral rule, the second line follows from plugging $\widehat{\lambda}(\gamma(\pi))$ in Equation (58) and using Equation (42), Equation (43) and Equation (59), while the inequality follows from the fact that each of the term within the square brackets is strictly positive.

Step 3. Let

$$\pi^\circ(b,\phi) \triangleq \left\{\pi \in \left(0, \left(\frac{\frac{1}{X}}{1-\frac{1}{X}}\right)^2\right) \middle| W_T(\pi) = W_O(\pi) \;\text{ when}\, \phi < 1-2b\right\} \tag{62}$$

be the π that solves $W_T(\pi) = W_O(\pi)$ where $W_T(\pi)$ is given by Equation (48) with $\widetilde{\lambda}(\pi) = \widehat{\lambda}(\gamma(\pi))$, while $W_O(\pi)$ is given by Equation (49) with $\eta(b,\phi) > 0$. $\pi^\circ(b,\phi)$ is increasing in ϕ and decreasing in b.

Let

$$\pi^\circ(b) \triangleq \left\{\pi \in \left(0, \left(\frac{\frac{1}{X}}{1-\frac{1}{X}}\right)^2\right) \middle| W_T(\pi) = W_O(\pi) \;\text{ when}\, \phi \geq 1-2b\right\} \tag{63}$$

be the π that solves $W_T(\pi) = W_O(\pi)$ where $W_T(\pi)$ is given by Equation (48) with $\widetilde{\lambda}(\pi) = \widehat{\lambda}(\gamma(\pi))$, while $W_O(\pi)$ is given by Equation (49) with $\eta(b,\phi) = 0$. $\pi^\circ(b)$ is decreasing in b.

From the shapes of $W_T(\pi)$ and $W_O(\pi)$ described in Step 2, it follows that:

- if $b \leq b^\circ(X)$ where $b^\circ(X)$ is defined in Equation (55), we have that

- if $\phi \in (0, \phi^\circ(b))$ with $\phi^\circ(b)$ defined in Equation (56), we have $W_O(\pi) > W_T(\pi)$ for every $\pi \in (0,1)$,
- if $\phi \in [\phi^\circ(b), 1-2b)$, we have that
  - $*$ $W_O(\pi) > W_T(\pi)$ for $\pi \in (\pi^\circ(b,\phi), 1)$, where $\pi^\circ(b,\phi)$ is defined in Equation (62),
  - $*$ $W_O(\pi) \leq W_T(\pi)$ for $\pi \in (0, \pi^\circ(b,\phi)]$,
- if $\phi \in [1-2b, 1)$, we have that
  - $*$ $W_O(\pi) = W_T(\pi)$ if $\pi \in \Big[\big(\frac{1}{X}/(1-\frac{1}{X})\big)^2, 1\Big)$,
  - $*$ $W_O(\pi) > W_T(\pi)$ if $\pi \in \Big(\pi^\circ(b), \big(\frac{1}{X}/(1-\frac{1}{X})\big)^2\Big)$, where $\pi^\circ(b)$ is defined in Equation (63),
  - $*$ $W_O(\pi) \leq W_T(\pi)$ if $\pi \in (0, \pi^\circ(b)(b)]$,

- if $b > b^\circ(X)$, we have that
  - if $\phi \in (0, 1-2b)$ we have $W_O(\pi) > W_T(\pi)$ for every $\pi \in (0,1)$,
  - if $\phi \in [1-2b, 1)$ we have that
    - $*$ $W_O(\pi) = W_T(\pi)$ if $\pi \in \Big[\big(\frac{1}{X}/(1-\frac{1}{X})\big)^2, 1\Big)$,
    - $*$ $W_O(\pi) > W_T(\pi)$ if $\pi \in \Big(0, \big(\frac{1}{X}/(1-\frac{1}{X})\big)^2\Big)$.

■

Proof of Proposition 5. We proceed in 4 steps: Step 1 derives the privacy-concerned borrowers' ex-ante surplus when the lender employs a transparent and an opaque algorithm, Step 2 compares their surplus across regimes, Step 3 derives the strategic borrowers' ex-ante surplus when the lender employs a transparent and an opaque algorithm, Step 4 compares their surplus across regimes.

Preliminaries. Consider the strategic borrower's data sharing strategy of the form

$$m_Q(z) = \begin{cases} \varnothing & \text{if } z \in Q \\ z & \text{if } z \notin Q, \end{cases}$$

where $Q \in \{\mathcal{G}(\lambda,\pi), \mathcal{H}(b,\phi)\}$ is the data-withholding set induced by a specific transparency regime, and where $\mathcal{G}(\lambda,\pi)$ and $\mathcal{H}(b,\phi)$ are defined in Proposition 1 and Proposition 2, respectively.

The privacy-concerned borrower's ex-ante equilibrium surplus from a transparency regime $\tau \in \{T, O\}$ inducing a data-withholding set Q is:

$$\begin{aligned} V_\tau^{B,p}(z) &\triangleq \frac{1}{2}\int_{-1}^{1} V_\lambda^B(\varnothing, z)\,\mathrm{d}\lambda, \\ &= \int_0^1 V_\lambda^B(\varnothing, z)\,\mathrm{d}\lambda, \end{aligned} \tag{64}$$

where $V_\lambda^B(\varnothing, z)$ is defined in Equation (24), while the second line of Equation (64) follows from the borrower's profit function being even in λ.

The strategic borrower's ex-ante equilibrium surplus from a transparency regime $\tau \in \{T, O\}$ inducing a data-withholding set Q is:

$$V_{\tau}^{B,s}(z) \triangleq \frac{1}{2}\int_{-1}^{1} V_{\lambda}^{B}(m_Q(z), z)\,\mathrm{d}\lambda, \tag{65}$$

$V_{\lambda}^{B}(z,z) = V_{\lambda}^{B}(z)$ is defined in Equation (20) while $V_{\lambda}^{B}(\varnothing, z)$ is defined in Equation (24).

Step 1. Proceeding as in Step 1 and Step 2 of Proposition 3, we get the privacy-concerned borrowers' surpluses from a transparent and an opaque algorithm. These are, respectively:

$$V_{T}^{B,p}(z) = \int_{0}^{\widetilde{\lambda}(\pi)} (1-\phi)\mu_{\lambda}(z)\left(X - \frac{1}{\mu_{\lambda}(\gamma(\pi))}\right) + b\ \mathrm{d}\lambda \tag{66}$$

and

$$V_{O}^{B,p}(z) = \int_{0}^{1} (1-\phi)\mu_{\lambda}(z)\left(X - \frac{1}{\mu_{\lambda}(\frac{1}{2})}\right) + b\ \mathrm{d}\lambda \tag{67}$$

Step 2. Note that the integrand in Equation (67) is strictly bigger than the integrand in Equation (66) since when $\lambda > 0$ we have $\mu_{\lambda}(\frac{1}{2}) > \mu_{\lambda}(\gamma(\pi))$ as $\gamma(\pi) < \frac{1}{2}$. Moreover, both integrands are strictly positive and $\widetilde{\lambda}(\pi) \leq 1$. It follows that $V_{O}^{B,p}(z) > V_{T}^{B,p}(z)$.

Step 3. We now compute the ex-ante surplus of strategic borrowers, conditional on z.

When the lender uses a transparent algorithm, the set of borrowers that withhold data is $Q = \mathcal{G}(\lambda, \pi)$, given in Proposition 1. We can distinguish three types of borrowers:

- Borrowers with $z \in [0, \gamma(\pi)]$ withhold information when $\lambda > 0$ and obtain credit only if $r(\lambda) < \gamma(\pi)$, i.e. if $\lambda < \widehat{\lambda}(\gamma(\pi))$, while they are denied credit when $\lambda \geq \widehat{\lambda}(\gamma(\pi))$. If $\gamma(\pi) > \frac{1}{X}$ they always obtain credit, i.e. $\widehat{\lambda}(\gamma(\pi)) > 1$. Moreover, they disclose information and obtain credit whenever $\lambda < 0$. It follows that their surplus conditional on λ is

$$V_{\lambda}^{B}(m_Q(z), z) = \begin{cases} (1-\phi)\mu_{\lambda}(z)\left(X - \dfrac{1}{\mu_{\lambda}(z)}\right) + b & \text{if } \lambda \in [-1, 0) \\ \mathbb{1}\{\lambda < \widetilde{\lambda}(\pi)\}\left[(1-\phi)\mu_{\lambda}(z)\left(X - \dfrac{1}{\mu_{\lambda}(\gamma(\pi))}\right) + b\right] & \text{if } \lambda \in [0, 1], \end{cases}$$

  where $\widetilde{\lambda}(\pi) = \min\{\widehat{\lambda}(\gamma(\pi)), 1\}$.

- Borrowers with $z \in [1-\gamma(\pi), 1]$ withhold information when $\lambda < 0$ and obtain credit only if $1-\gamma(\pi) < r(\lambda)$, i.e. if $\lambda > -\widehat{\lambda}(\gamma(\pi))$, while they are denied credit when $\lambda \leq -\widehat{\lambda}(\gamma(\pi))$. If $1-\gamma(\pi) < 1 - \frac{1}{X}$ they always obtain credit and $-\widehat{\lambda}(\gamma(\pi)) < -1$. Moreover, they disclose information and obtain credit whenever $\lambda > 0$. It follows that their surplus conditional on λ is

$$V_{\lambda}^{B}(m_Q(z), z) = \begin{cases} \mathbb{1}\{\lambda > -\widetilde{\lambda}(\pi)\}\left[(1-\phi)\mu_{\lambda}(z)\left(X - \dfrac{1}{\mu_{\lambda}(1-\gamma(\pi))}\right) + b\right] & \text{if } \lambda \in [-1, 0) \\ (1-\phi)\mu_{\lambda}(z)\left(X - \dfrac{1}{\mu_{\lambda}(z)}\right) + b & \text{if } \lambda \in [0, 1]. \end{cases}$$

- Borrowers with $z \in (\gamma(\pi), 1-\gamma(\pi))$ disclose information and obtain credit only if $z > r(\lambda)$ when $\lambda > 0$ and when $z < r(\lambda)$ when $\lambda < 0$, that is when $\lambda \in (\lambda_-(z), \lambda_+(z))$, where

$$\lambda_-(z) \triangleq \begin{cases} -1 & \text{if } z \in \left[0, 1-\dfrac{1}{X}\right] \\ \widehat{\lambda}(z) & \text{if } z \in \left(1-\dfrac{1}{X}, 1\right], \end{cases} \qquad \lambda_+(z) \triangleq \begin{cases} \widehat{\lambda}(z) & \text{if } z \in \left[0, \dfrac{1}{X}\right) \\ 1 & \text{if } z \in \left[\dfrac{1}{X}, 1\right]. \end{cases} \tag{68}$$

It follows that their surplus conditional on λ is

$$V_\lambda^B(m_Q(z), z) = \mathbb{1}\{\lambda \in (\lambda_-(z), \lambda_+(z))\}\left[(1-\phi)\mu_\lambda(z)\left(X - \frac{1}{\mu_\lambda(z)}\right) + b\right].$$

Integrating over λ as in Equation (65), we can write the strategic borrowers' ex-ante surplus under transparency as

$$\begin{aligned} V_T^{B,s}(z) &= \begin{cases} \displaystyle\int_{-1}^{0}\left[\mu_\lambda(z)(1-\phi)\left(X-\frac{1}{\mu_\lambda(z)}\right)+b\right]\frac{1}{2}\,\mathrm{d}\lambda + \\ \qquad + \displaystyle\int_{0}^{\widetilde{\lambda}(\pi)}\left[\mu_\lambda(z)(1-\phi)\left(X-\frac{1}{\mu_\lambda(\gamma(\pi))}\right)+b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \big[0, \gamma(\pi)\big] \\ \displaystyle\int_{\lambda_-(z)}^{\lambda_+(z)}\left[\mu_\lambda(z)(1-\phi)\left(X-\frac{1}{\mu_\lambda(z)}\right)+b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \big(\gamma(\pi), 1-\gamma(\pi)\big) \\ \displaystyle\int_{-\widetilde{\lambda}(\pi)}^{0}\left[\mu_\lambda(z)(1-\phi)\left(X-\frac{1}{\mu_\lambda(1-\gamma(\pi))}\right)+b\right]\frac{1}{2}\,\mathrm{d}\lambda + \\ \qquad + \displaystyle\int_{0}^{1}\left[\mu_\lambda(z)(1-\phi)\left(X-\frac{1}{\mu_\lambda(z)}\right)+b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \big[1-\gamma(\pi), 1\big], \end{cases} \\ &= V^B(z) + O^B(z,\pi) \geq V^B(z), \end{aligned} \tag{69}$$

where $V^B(z)$ is the expected borrower's surplus from full disclosure and is defined in Equation (28) while

$$O^B(z,\pi) \triangleq \begin{cases} \displaystyle\int_0^{\widehat{\lambda}(z)} \mu_\lambda(z)(1-\phi)\left(\frac{1}{\mu_\lambda(z)} - \frac{1}{\mu_\lambda(\gamma(\pi))}\right)\frac{1}{2}\,\mathrm{d}\lambda + \\ \qquad + \displaystyle\int_{\widehat{\lambda}(z)}^{\widetilde{\lambda}(\pi)} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(\gamma(\pi))}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \big[0, \gamma(\pi)\big] \\ 0 & \text{if } z \in \big(\gamma(\pi), 1-\gamma(\pi)\big) \\ \displaystyle\int_{-\widetilde{\lambda}(\pi)}^{\widehat{\lambda}(z)} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(1-\gamma(\pi))}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda + \\ \qquad + \displaystyle\int_{\widehat{\lambda}(z)}^{0} \mu_\lambda(z)(1-\phi)\left(\frac{1}{\mu_\lambda(z)} - \frac{1}{\mu_\lambda(1-\gamma(\pi))}\right)\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \big[1-\gamma(\pi), 1\big], \end{cases} \tag{70}$$

with $O^B(z,\pi) > 0$ for $z \in [0,\gamma(\pi)) \cup (1-\gamma(\pi),1]$ since every integrand is positive as we have that i) $\mu_\lambda(\gamma(\pi)) > \mu_\lambda(z)$ for $\lambda > 0$, which also implies that $\mu_\lambda(\gamma(\pi))X - 1 > \mu_\lambda(z)X - 1 > 0$ for $\lambda > \widehat{\lambda}(z)$, and ii) $\mu_\lambda(1-\gamma(\pi)) > \mu_\lambda(z)$ for $\lambda < 1$, which also implies that $\mu_\lambda(1-\gamma(\pi))X - 1 > \mu_\lambda(z)X - 1 > 0$ for $\lambda < -\widehat{\lambda}(z)$.

When the lender uses an opaque algorithm, the set of borrowers that withhold data is $Q = \mathcal{H}(b\,\phi)$, given in Proposition 2. Borrowers with $z \in [0, \eta(b,\phi))$ share data and obtain credit only if $\lambda < \widehat{\lambda}(z)$, borrowers with $z \in (1-\eta(b,\phi), 1]$ share data and obtain credit only if $\lambda > \widehat{\lambda}(z)$, instead borrowers with $z \in [\eta(b,\phi), 1-\eta(b,\phi)]$ withhold data and always obtain credit. Integrating over λ as in Equation (65), we can write the strategic borrowers' surplus under opacity as

$$\begin{aligned} V_O^{B,s}(z) &= \begin{cases} \displaystyle\int_{-1}^{\widehat{\lambda}(z)} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(z)}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \big[0, \eta(b,\phi)\big) \\ \displaystyle\int_{-1}^{1} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\frac{1}{2}}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \big[\eta(b,\phi), 1-\eta(b,\phi)\big] \\ \displaystyle\int_{\widehat{\lambda}(z)}^{1} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(z)}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \big(1-\eta(b,\phi), 1\big], \end{cases} \\ &= \max\left\{V^B(z), V^B(\varnothing)\right\} \end{aligned} \tag{71}$$

where $V^B(z)$ is the expected borrower's surplus from full disclosure defined in Equation (28) while $V^B(\varnothing)$ is the expected borrower's surplus from no disclosure defined in Equation (31).

Step 4. Let

$$\Delta^{B,s}(z) \triangleq V_T^{B,s}(z) - V_O^{B,s}(z)$$

be the difference between the strategic borrower's surplus with a transparent and an opaque algorithm. We now study the sign of $\Delta^{B,s}(z)$, for $z \leq \frac{1}{2}$. The analysis is symmetric for $z \geq \frac{1}{2}$. We distinguish 2 cases.

Case 1 ($\phi \geq 1-2b$). For $\phi \geq 1-2b$ we have $\eta(b,\phi) = 0$.

Consider first $\pi \geq \widehat{\pi}(X)$ where

$$\widehat{\pi}(X) \triangleq \left(\frac{\frac{1}{X}}{1-\frac{1}{X}}\right)^2 \in (0,1). \tag{72}$$

For these parameter values we have $\gamma(\pi) \geq \frac{1}{X}$, so $\widetilde{\lambda}(\pi) = 1$ and $\lambda_-(z) = -1$ for all $z > \gamma(\pi)$ and $\lambda_+(z) = 1$ for all $z < 1 - \gamma(\pi)$. Noting that the second piece of Equation (71) can be rewritten as

$$\int_{-1}^{1} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\frac{1}{2}}\right) + b\right] \frac{1}{2}\,\mathrm{d}\lambda = \int_{-1}^{1} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(z)}\right) + b\right] \frac{1}{2}\,\mathrm{d}\lambda \tag{73}$$

we have

$$\Delta^{B,s}(z) = \begin{cases} \displaystyle\int_0^1 \mu_\lambda(z)(1-\phi)\left(\frac{1}{\mu_\lambda(z)} - \frac{1}{\mu_\lambda(\gamma(\pi))}\right)\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in [0, \gamma(\pi)] \\ 0 & \text{if } z \in (\gamma(\pi), \frac{1}{2}]. \end{cases} \tag{74}$$

This is weakly positive and strictly so for $z \in [0, \gamma(\pi))$ since $\mu_\lambda(\gamma(\pi)) > \mu_\lambda(z)$ for $\lambda > 0$.

Consider now $\pi < \widehat{\pi}(X)$, that is $\gamma(\pi) < \frac{1}{X}$ so that $\widetilde{\lambda}(\pi) = \widehat{\lambda}(\gamma(\pi)) < 1$ and $\lambda_+(z) \leq 1$ and $\lambda_-(z) \geq -1$. Using again Equation (73) we get

$$\Delta^{B,s}(z) = \begin{cases} \displaystyle\int_0^{\widehat{\lambda}(\gamma(\pi))} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(\gamma(\pi))}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda + \\ \qquad\qquad - \displaystyle\int_0^1 \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(z)}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in [0, \gamma(\pi)] \\ - \displaystyle\int_{\widehat{\lambda}(z)}^1 \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(z)}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \left(\gamma(\pi), \frac{1}{X}\right) \\ 0 & \text{if } z \in \left[\frac{1}{X}, \frac{1}{2}\right]. \end{cases} \tag{75}$$

The sign of $\Delta^{B,s}(z)$ depends on the value of π and z.

First, for $z \in (\gamma(\pi), \frac{1}{X})$ we have $\Delta^{B,s}(z) < 0$. To see this note that for $z \in (\gamma(\pi), \frac{1}{X})$ we have $V_T^{B,s}(z) = V^B(z)$ where $V^B(z)$ is the borrowers' expected utility from data sharing under an opaque algorithm and is defined in Equation (28) while $V_O^{B,s}(z) = V^B(\varnothing)$ where $V^B(\varnothing)$ is the borrowers' utility from data withholding under an opaque algorithm defined in Equation (31). It follows that $\Delta^{B,s}(z) = V^B(z) - V^B(\varnothing) < 0$ since for $\phi \geq 1 - 2b$ it is strictly optimal to withhold data for $z \in [0, \frac{1}{X})$ as shown in Proposition 2.

Second, for $z \in [0, \gamma(\pi)]$ we have that $\Delta^{B,s}(z)$ is strictly decreasing in z since its first-order derivative on that interval is

$$\frac{\partial \Delta^{B,s}}{\partial z} = \frac{1}{2}(1-\phi)X\left[\left(\int_0^{\widehat{\lambda}(\gamma(\pi))} \lambda \mathrm{d}\lambda - \int_0^1 \lambda \mathrm{d}\lambda\right) - \int_0^{\widehat{\lambda}(\gamma(\pi))} \lambda \frac{1}{X}\frac{1}{\mu_\lambda(\gamma(\pi))}\mathrm{d}\lambda\right] < 0, \tag{76}$$

since the term in brackets is negative as $\widehat{\lambda}(\gamma(\pi)) < 1$ while the integrand in the second addend is strictly positive.

Third, for $z \in [0, \gamma(\pi)]$ we have that $\Delta^{B,s}(z)$ is strictly increasing in π since

$$\begin{aligned}\frac{\partial \Delta^{B,s}}{\partial \pi} &= \widehat{\lambda}'(\gamma(\pi))\gamma'(\pi)\left[\mu_{\widehat{\lambda}(\gamma(\pi))}(z)(1-\phi)\left(X - \frac{1}{\mu_{\widehat{\lambda}(\gamma(\pi))}}\right) + b\right]\frac{1}{2} + \\ &\quad + \int_0^{\widehat{\lambda}(\gamma(\pi))}(1-\phi)\frac{\mu_\lambda(z)}{\big(\mu_\lambda(\gamma(\pi))\big)^2}\lambda\gamma'(\pi)\,\mathrm{d}\lambda \\ &= \widehat{\lambda}'(\gamma(\pi))\gamma'(\pi)\frac{1}{2}b + \int_0^{\widehat{\lambda}(\gamma(\pi))}(1-\phi)\frac{\mu_\lambda(z)}{\big(\mu_\lambda(\gamma(\pi))\big)^2}\lambda\gamma'(\pi)\,\mathrm{d}\lambda > 0,\end{aligned} \tag{77}$$

where the first and second lines use Leibnitz integral rule, the last line follows from the fact that

$$\mu_{\widehat{\lambda}(\gamma(\pi))} = \frac{1}{2} + \frac{\frac{1}{2} - \frac{1}{X}}{\frac{1}{2} - \gamma(\pi)}\left(\gamma(\pi) - \frac{1}{2}\right) = \frac{1}{X},$$

while the inequality follows from Equation (59) and the fact that the integrand is positive since $\gamma'(\pi) > 0$.

Forth, since we are considering $\pi \in (0, \widehat{\pi}(X))$ we can evaluate $\Delta^{B,s}(0)$ at its extreme values of π to get

$$\begin{aligned}\lim_{\pi\to 0}\Delta^{B,s}(0) &= -\int_{\overline{\lambda}}^1\Big[(1-\phi)\big(\mu_\lambda(0)X - 1\big) + b\Big]\frac{1}{2}\,\mathrm{d}\lambda < 0 \\ \lim_{\pi\to\left(\frac{1}{X}/\left(1-\frac{1}{X}\right)\right)^2}\Delta^{B,s}(0) &= \int_0^1 \mu_\lambda(0)(1-\phi)\left(\frac{1}{\mu_\lambda(0)} - \frac{1}{\mu_\lambda(\frac{1}{X})}\right)\frac{1}{2}\,\mathrm{d}\lambda > 0.\end{aligned} \tag{78}$$

Since $\Delta^{B,s}(0)$ is strictly increasing in π (see Equation (77)), there exists a threshold $\pi^*(b,\phi) \in (0, \widehat{\pi}(X))$ such that $\Delta^{B,s}(0) < 0$ if and only if $\pi < \pi^*(b,\phi)$. Let this threshold be defined as

$$\pi^*(b,\phi) \triangleq \left\{\pi \in \big(0, \widehat{\pi}(X)\big)\,\middle|\,\Delta^{B,s}(0) = 0\right\}. \tag{79}$$

Since $\Delta^{B,s}(z)$ is strictly decreasing in z (see Equation (76)), if $\pi < \pi^*(b,\phi)$ then $\Delta^{B,s}(z) < 0$ for every $z \in [0, \gamma(\pi)]$. Instead, since $\Delta^{B,s}(z)$ is strictly decreasing in z for $z \in [0, \gamma(\pi)]$ and strictly increasing in π, when $\pi \geq \pi^*(b,\phi)$, there exists an increasing threshold $z^*(\pi)$ such that $\Delta^{B,s}(z) < 0$ if and only if $z > z^*(\pi)$. Let this threshold be defined as

$$z^*(\pi) \triangleq \left\{z \in (0, \gamma(\pi))\,\middle|\,\Delta^{B,s}(z) = 0 \text{ for } z \in [0, \gamma(\pi)]\right\}. \tag{80}$$

In summary, by symmetry of $\Delta^{B,s}(z)$ around $z = \frac{1}{2}$, when $\phi \geq 1 - 2b$ we have that:

- if $\pi \in (0, \pi^*(b,\phi))$, where $\pi^*(b,\phi)$ is defined in Equation (79), we have
  - $V_T^{B,s}(z) < V_O^{B,s}(z)$ for $z \in [0, \frac{1}{X}) \cup (1 - \frac{1}{X}, 1]$, and

- $V_T^{B,s}(z) = V_O^{B,s}(z)$ for $z \in [\frac{1}{X}, 1 - \frac{1}{X}]$;

- if $\pi \in [\pi^*(b,\phi), \widehat{\pi}(X))$, where $\widehat{\pi}(X)$ is defined in Equation (72), we have
  - $V_T^{B,s}(z) > V_O^{B,s}(z)$ for $z \in [0, z^*(\pi)) \cup (1 - z^*(\pi), 1]$, where $z^*(\pi)$ is defined in Equation (80), and
  - $V_T^{B,s}(z) < V_O^{B,s}(z)$ for $z \in (z^*(\pi), \frac{1}{X}) \cup (1 - z^*(\pi), 1 - \frac{1}{X})$, and
  - $V_T^{B,s}(z) = V_O^{B,s}(z)$ for $z \in [\frac{1}{X}, 1 - \frac{1}{X}] \cup \{z^*(\pi), 1 - z^*(\pi)\}$,
- if $\pi \in [\widehat{\pi}(X), 1)$ we have
  - $V_T^{B,s}(z) > V_O^{B,s}(z)$ for $z \in [0, \gamma(\pi)) \cup (1 - \gamma(\pi), 1]$, and
  - $V_T^{B,s}(z) = V_O^{B,s}(z)$ for $z \in [\gamma(\pi), 1 - \gamma(\pi)]$.

Case 2 ($\phi < 1 - 2b$). For $\phi < 1 - 2b$ we have $\eta(b,\phi) > 0$.

Consider $\pi \geq \widehat{\pi}(X)$ where $\widehat{\pi}(X)$ is defined in Equation (72), so that we have $\gamma(\pi) \geq \frac{1}{X}$ and $\widetilde{\lambda}(\pi) = 1$ and $\lambda_-(z) = -1$ for all $z > \gamma(\pi)$ and $\lambda_+(z) = 1$ for all $z < 1 - \gamma(\pi)$. Using Equation (73), we have

$$\Delta^{B,s}(z) = \begin{cases} \displaystyle\int_0^{\widehat{\lambda}(z)} \mu_\lambda(z)(1-\phi)\left(\frac{1}{\mu_\lambda(z)} - \frac{1}{\mu_\lambda(\gamma(\pi))}\right)\frac{1}{2}\,\mathrm{d}\lambda \; + & \\ \qquad\qquad \displaystyle + \int_{\widehat{\lambda}(z)}^{\widetilde{\lambda}(\pi)} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(\gamma(\pi))}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in [0, \eta(b,\phi)) \\ \displaystyle\int_0^1 \mu_\lambda(z)(1-\phi)\left(\frac{1}{\mu_\lambda(z)} - \frac{1}{\mu_\lambda(\gamma(\pi))}\right)\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in [\eta(b,\phi), \gamma(\pi)] \\ 0 & \text{if } z \in (\gamma(\pi), \frac{1}{2}). \end{cases} \tag{81}$$

This is weakly positive and strictly so for $z \in [0, \gamma(\pi))$. To see this note that this $\Delta^{B,s}(z)$ is the same as in Equation (74) for $z \geq \eta(b,\phi)$, and we already prove that $\Delta^{B,s}(z) > 0$ for $z \in [\eta(b,\phi), \gamma(\pi))$ after Equation (74). For $z \in [0, \eta(b,\phi))$ note that $\Delta^{B,s}(z) = O^{B,s}(z) > 0$, where $O^{B,s}(z)$ is defined in Equation (70).

Consider now $\pi \in [\pi^{**}(b,\phi), \widehat{\pi}(X))$ where

$$\pi^{**}(b,\phi) \triangleq \left(\frac{\eta(b,\phi)}{1 - \eta(b,\phi)}\right)^2 \in (0, \widehat{\pi}(X)). \tag{82}$$

For these parameter values we have $\eta(b,\phi) \leq \gamma(\pi) < \frac{1}{X}$ so that $\widetilde{\lambda}(\pi) = \widehat{\lambda}(\gamma(\pi)) < 1$ and $\lambda_+(z) \leq 1$ and

$\lambda_-(z) \geq -1$. Using again Equation (73) we get

$$\Delta^{B,s}(z) = \begin{cases} \displaystyle\int_0^{\widehat{\lambda}(z)} \mu_\lambda(z)(1-\phi)\left(\frac{1}{\mu_\lambda(z)} - \frac{1}{\mu_\lambda(\gamma(\pi))}\right)\frac{1}{2}\,\mathrm{d}\lambda + \\ \qquad + \displaystyle\int_{\widehat{\lambda}(z)}^{\widetilde{\lambda}(\pi)} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(\gamma(\pi))}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in [0, \eta(b,\phi)) \\ \displaystyle\int_0^{\widehat{\lambda}(\gamma(\pi))} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(\gamma(\pi))}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda + \\ \qquad - \displaystyle\int_0^1 \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(z)}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in [\eta(b,\phi), \gamma(\pi)] \\ -\displaystyle\int_{\widehat{\lambda}(z)}^1 \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(z)}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \left(\gamma(\pi), \frac{1}{X}\right) \\ 0 & \text{if } z \in \left[\frac{1}{X}, \frac{1}{2}\right]. \end{cases} \tag{83}$$

Note that this $\Delta^{B,s}(z)$ is the same as in Equation (75) for $z \geq \eta(b,\phi)$. We already proved that $\Delta^{B,s}(z) < 0$ for $z \in (\gamma(\pi), \frac{1}{X})$ (see the discussion following Equation (75)). For $z \in [0, \eta(b,\phi))$ we have $\Delta^{B,s}(z) = O^{B,s}(z) > 0$ where $O^{B,s}(z)$ is defined in Equation (70). Since $\Delta^{B,s}(z)$ is continuous (as it is the difference of two continuous functions), decreasing in z for $z \in [\eta(b,\phi), \gamma(\pi)]$ (see Equation (76) for $z \in [0, \gamma(\pi)]$) and increasing in π for $z \in [\eta(b,\phi), \gamma(\pi)]$ (see Equation (77)) there exists an increasing $z^*(\pi) \in [\eta(b,\phi), \gamma(\pi)]$ defined in Equation (80) such that $\Delta^{B,s}(z) < 0$ if and only if $z > z^*(\pi)$.

Consider now $\pi < \pi^{**}(b,\phi)$ so that $\gamma(\pi) < \eta(b,\phi)$. Using again Equation (73) we get

$$\Delta^{B,s}(z) = \begin{cases} \displaystyle\int_0^{\widehat{\lambda}(z)} \mu_\lambda(z)(1-\phi)\left(\frac{1}{\mu_\lambda(z)} - \frac{1}{\mu_\lambda(\gamma(\pi))}\right)\frac{1}{2}\,\mathrm{d}\lambda + \\ \qquad + \displaystyle\int_{\widehat{\lambda}(z)}^{\widetilde{\lambda}(\pi)} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(\gamma(\pi))}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in [0, \gamma(\pi)] \\ 0 & \text{if } z \in (\gamma(\pi), \eta(b,\phi)) \\ -\displaystyle\int_{\widehat{\lambda}(z)}^1 \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(z)}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \left[\eta(b,\phi), \frac{1}{X}\right) \\ 0 & \text{if } z \in \left[\frac{1}{X}, \frac{1}{2}\right]. \end{cases} \tag{84}$$

Once again $\Delta^{B,s}(z) = O^{B,s}(z) > 0$ for $z \in [0, \gamma(\pi)]$, where $O^{B,s}(z)$ is defined in Equation (70). Moreover, $\Delta^{B,s}(z) < 0$ for $z \in [\eta(b,\phi), \frac{1}{X})$, since in this interval we have $V_T^{B,s}(z) = V^B(z)$ where $V^B(z)$ is the borrowers' expected utility from data sharing under an opaque algorithm and is defined in Equation (28) while $V_O^{B,s}(z) = V^B(\varnothing)$ where $V^B(\varnothing)$ is the borrowers' utility from data withholding under an opaque algorithm defined in Equation (31). It follows that $\Delta^{B,s}(z) = V^B(z) - V^B(\varnothing) < 0$ since for $z \in [\eta(b,\phi), \frac{1}{X})$ it is strictly optimal to withhold data as shown in Proposition 2.

In summary, by symmetry of $\Delta^{B,s}(z)$ around $z = \frac{1}{2}$, when $\phi < 1 - 2b$ we have that:

- if $\pi \in (0, \pi^{**}(b,\phi))$, where $\pi^{**}(b,\phi)$ is defined in Equation (82), we have
  - $V_T^{B,s}(z) > V_O^{B,s}(z)$ for $z \in [0, \gamma(\pi)) \cup (1-\gamma(\pi), 1]$, and
  - $V_T^{B,s}(z) = V_O^{B,s}(z)$ for $z \in [\gamma(\pi), \eta(b,\phi)) \cup [\frac{1}{X}, 1 - \frac{1}{X}] \cup (1-\eta(b,\phi), 1-\gamma(\pi)]$, and
  - $V_T^{B,s}(z) < V_O^{B,s}(z)$ for $z \in [\eta(b,\phi), \frac{1}{X}) \cup (1 - \frac{1}{X}, 1-\eta(b,\phi)]$;
- if $\pi \in [\pi^{**}(b,\phi), \widehat{\pi}(X))$, where $\widehat{\pi}(X)$ is defined in Equation (72), we have
  - $V_T^{B,s}(z) > V_O^{B,s}(z)$ for $z \in [0, z^*(\pi)) \cup (1 - z^*(\pi), 1]$, where $z^*(\pi)$ is defined in Equation (80), and
  - $V_T^{B,s}(z) < V_O^{B,s}(z)$ for $z \in (z^*(\pi), \frac{1}{X}) \cup (1 - z^*(\pi), 1 - \frac{1}{X})$, and
  - $V_T^{B,s}(z) = V_O^{B,s}(z)$ for $z \in [\frac{1}{X}, 1 - \frac{1}{X}] \cup \{z^*(\pi), 1 - z^*(\pi)\}$,
- if $\pi \in [\widehat{\pi}(X), 1)$ we have
  - $V_T^{B,s}(z) > V_O^{B,s}(z)$ for $z \in [0, \gamma(\pi)) \cup (1-\gamma(\pi), 1]$, and
  - $V_T^{B,s}(z) = V_O^{B,s}(z)$ for $z \in [\gamma(\pi), 1-\gamma(\pi)]$.

Considering both Case 1 ($\phi \geq 1 - 2b$) and Case 2 ($\phi < 1 - 2b$) we can determine a threshold $z^\star(\pi)$ such that the surplus of a strategic borrower with data z is strictly higher under transparency if and only if $z \in [0, z^\star(\pi)) \cup (1 - z^\star(\pi), 1]$. This threshold is

$$z^\star(\pi) \triangleq \begin{cases} \begin{cases} 0 & \text{if } \pi \in \big(0, \pi^*(b,\phi)\big) \\ z^*(\pi) & \text{if } \pi \in \big[\pi^*(b,\phi), \widehat{\pi}(X)\big) \\ \gamma(\pi) & \text{if } \pi \in \big[\widehat{\pi}(X), 1\big) \end{cases} & \text{if } \phi \geq 1 - 2b, \\ \begin{cases} \gamma(\pi) & \text{if } \big(0, \pi^{**}(b,\phi)\big) \\ z^*(\pi) & \text{if } \big[\pi^{**}(b,\phi), \widehat{\pi}(X)\big) \\ \gamma(\pi) & \text{if } \big[\widehat{\pi}(X), 1\big) \end{cases} & \text{if } \phi < 1 - 2b, \end{cases} \tag{85}$$

where $\pi^*(b,\phi)$, $\widehat{\pi}(X)$ and $\pi^{**}(b,\phi)$ are defined in Equations (79), (72) and (82), respectively, while $z^*(\pi)$ and $\gamma(\pi)$ are defined in Equations (80) and (12). ■

Proof of Proposition 6. We proceed in 3 steps: Step 1 derives the borrower's surplus when the lender employs a transparent and an opaque algorithm, Step 2 studies the graph of the borrower's surplus as a function of π; Step 3 compares the graphs.

Preliminaries. Consider a data sharing strategy of the form

$$m_Q(z) = \begin{cases} \varnothing & \text{if } z \in Q \\ z & \text{if } z \notin Q, \end{cases}$$

where $Q \in \{\mathcal{G}(\lambda,\pi), \mathcal{H}(b,\phi)\}$ is the data-withholding set induced by a specific transparency regime, and where $\mathcal{G}(\lambda,\pi)$ and $\mathcal{H}(b,\phi)$ are defined in Proposition 1 and Proposition 2, respectively. The borrower's equilibrium surplus from a transparency regime $\tau \in \{T, O\}$ inducing a data-withholding set Q is:

$$\begin{aligned} V_\tau^B(\pi) &\triangleq \frac{1}{2}\int_{-1}^{1} V_Q^B(\pi,\lambda)\,\mathrm{d}\lambda, \\ &= \int_0^1 V_Q^B(\pi,\lambda)\,\mathrm{d}\lambda, \end{aligned} \tag{86}$$

where

$$V_Q^B(\pi,\lambda) \triangleq \pi V_\lambda^B(\varnothing) + (1-\pi)\int_0^1 V_\lambda^B\big(m_Q(z)\big)\mathrm{d}z, \tag{87}$$

$V_\lambda^B(z)$ and $V_\lambda^B(\varnothing)$ are defined in Equation (20) and Equation (23), while the second line of Equation (86) follows from the borrower's profit function being even in λ.

Step 1. Proceeding as in Step 1 and Step 2 of Proposition 3, we get the borrower's surpluses from a transparent and an opaque algorithm. These are, respectively:

$$\begin{aligned} V_T^B(\pi) = \widetilde{\lambda}(\pi)\ &\left[(1-\phi)\left(\tfrac{1}{2}X - 1\right) + b\right] + \\ &+ (1-\pi)\int_{\widetilde{\lambda}(\pi)}^1 \int_{r(\lambda)}^1 (1-\phi)\big(\mu_\lambda(z)X - 1\big) + b\ \ \mathrm{d}z\,\mathrm{d}\lambda \end{aligned} \tag{88}$$

and

$$\begin{aligned} V_O^B(\pi) &= \overline{\lambda}\ \left[(1-\phi)\left(\tfrac{1}{2}X - 1\right) + b\right] + \\ &+ \int_{\overline{\lambda}}^{\widehat{\lambda}(\eta(b,\phi))} \pi\left[(1-\phi)\left(\tfrac{1}{2}X - 1\right) + b\right] + (1-\pi)\int_{r(\lambda)}^1 (1-\phi)\big(\mu_\lambda(z)X - 1\big) + b\ \ \mathrm{d}z\,\mathrm{d}\lambda + \\ &+ \int_{\widehat{\lambda}(\eta(b,\phi))}^1 \pi\left[(1-\phi)\left(\tfrac{1}{2}X - 1\right) + b\right] + (1-\pi)\int_{\eta(b,\phi)}^1 (1-\phi)\big(\mu_\lambda(z)X - 1\big) + b\ \ \mathrm{d}z\,\mathrm{d}\lambda. \end{aligned} \tag{89}$$

Step 2. We now study the graphs of the borrower's surplus in Equation (88) and Equation (89). For an arbitrary function $y : (0,1) \to \mathbb{R}$, we will use the simplified notation $y(0) = \lim_{\pi\to 0} y(\pi)$ and $y(1) = \lim_{\pi\to 1} y(\pi)$ to denote the limiting values of $y(\pi)$ as π approaches 0 and 1, respectively.

First, since $\widetilde{\lambda}(1) = 1$ and $\widetilde{\lambda}(0) = \overline{\lambda}$, we have that

$$V_O^B(1) = V_T^B(1) = (1-\phi)\left(\tfrac{1}{2}X - 1\right) + b,$$

and

$$
\begin{aligned}
V_O^B(0) - V_T^B(0) &= \int_{\widehat{\lambda}\left(\eta(b,\phi)\right)}^{1} \int_{\eta(b,\phi)}^{r(\lambda)} \left[(1-\phi)\left(\mu_\lambda(z)X - 1\right) + b\right] \mathrm{d}z\, \mathrm{d}\lambda, \\
&= \int_{\widehat{\lambda}\left(\eta(b,\phi)\right)}^{1} \left(r(\lambda) - \eta(b,\phi)\right) \left[(1-\phi)\left(\mu_\lambda\left(\tfrac{\eta(b,\phi)+r(\lambda)}{2}\right) X - 1\right) + b\right] \mathrm{d}\lambda > 0,
\end{aligned} \tag{90}
$$

where $\mu_\lambda(\cdot)$ is defined in Equation (2) and the first lines uses Equation (39). To prove the inequality in Equation (90), notice that $\widehat{\lambda}\left(\eta(b,\phi)\right) < 1$ for $b > 0$ and the integrand in the last line is positive since $r(\lambda) \geq \eta(b,\phi)$ for $\lambda \geq \widehat{\lambda}\left(\eta(b,\phi)\right)$ and the term in square brackets is positive if

$$
\frac{\eta(b,\phi) + r(\lambda)}{2} \geq \frac{1}{2} - \frac{1}{\lambda}\left(\frac{1}{2} - \frac{1}{X}\right) - \frac{1}{\lambda}\frac{1}{X}\frac{b}{1-\phi} = r(\lambda) - \frac{1}{\lambda}\frac{1}{X}\frac{b}{1-\phi},
$$

which reduces to $\lambda \leq 1$ when $\eta(b,\phi) > 0$ (i.e. when $\frac{1}{X}\left(1 - 2b/(1-\phi)\right) > 0$) and to $\lambda \leq 1 - 2\frac{1}{X}\left(1 - \frac{2b}{1-\phi}\right)$ when $\eta(b,\phi) = 0$ (i.e. when $\frac{1}{X}\left(1 - 2b/(1-\phi)\right) \leq 0$), and are both satisfied.

Second, notice that $V_O^B(\pi)$ is flat and equal to $(1-\phi)\left(\frac{1}{2}X - 1\right) + b = V_O^B(1)$ if $\phi \geq 1 - 2b$ (i.e. when $\eta(b,\phi) = 0$ and $\widehat{\lambda}(\eta(b,\phi)) = \overline{\lambda}$) and is strictly decreasing if $\phi < 1 - 2b$ (i.e. when $\eta(b,\phi) > 0$ and $\widehat{\lambda}(\eta(b,\phi)) > \overline{\lambda}$), since

$$
\begin{aligned}
\frac{\partial V_O^B}{\partial \pi} &= \int_{\overline{\lambda}}^{\widehat{\lambda}(\eta(b,\phi))} \left[(1-\phi)\left(\tfrac{1}{2}X - 1\right) + b\right] - \int_{r(\lambda)}^{1} (1-\phi)\left(\mu_\lambda(z)X - 1\right) + b \;\mathrm{d}z\,\mathrm{d}\lambda + \\
&+ \int_{\widehat{\lambda}(\eta(b,\phi))}^{1} \left[(1-\phi)\left(\tfrac{1}{2}X - 1\right) + b\right] - \int_{\eta(b,\phi)}^{1} (1-\phi)\left(\mu_\lambda(z)X - 1\right) + b \;\mathrm{d}z\,\mathrm{d}\lambda, \\
&= \int_{\overline{\lambda}}^{1} \int_{0}^{\min\{r(\lambda),\eta(b,\phi)\}} \left[(1-\phi)\left(\mu_\lambda(z)X - 1\right) + b\right] \mathrm{d}z\,\mathrm{d}\lambda, \\
&< \int_{\overline{\lambda}}^{1} \int_{0}^{\eta(b,\phi)} \left[(1-\phi)\left(\mu_\lambda(z)X - 1\right) + b\right] \mathrm{d}z\,\mathrm{d}\lambda, \\
&= (1-\overline{\lambda})\eta(b,\phi) \left[(1-\phi)\left(\left(\frac{\overline{\lambda}+1}{2}\frac{\eta(b,\phi)}{2} + \left(1 - \frac{\overline{\lambda}+1}{2}\right)\frac{1}{2}\right) X - 1\right) + b\right] < 0,
\end{aligned} \tag{91}
$$

where the third line follows from Equation (39) and the fact that $r(\lambda) \geq \eta(b,\phi)$ if and only if $\lambda \geq \widehat{\lambda}(\eta(b,\phi))$, while the last inequality reduces to $\phi < 1 - 2b$, the parameter values considered.

Third, $V_T^B(\pi)$ is flat and equal to $(1-\phi)\left(\frac{1}{2}X - 1\right) + b = V_T^B(1)$ if $\pi \geq \left(\frac{1}{X}/(1 - \frac{1}{X})\right)^2 \in (0,1)$ (i.e. when $\widetilde{\lambda}(\pi) = 1$) and is strictly convex for $\pi < \left(\frac{1}{X}/(1 - \frac{1}{X})\right)^2$ (i.e. when $\widetilde{\lambda}(\pi) = \widehat{\lambda}(\gamma(\pi)) < 1$) approaching the flat part from below. To prove that $V_T^B(\pi)$ is convex and approaches the flat part from below, rewrite it as

$$
V_T^B(\pi) = \widehat{\lambda}(\gamma(\pi)) \left[(1-\phi)\left(\tfrac{1}{2}X - 1\right) + b\right] + (1-\pi)\int_{\widehat{\lambda}(\gamma(\pi))}^{1} G(\lambda)\,\mathrm{d}\lambda
$$

where

$$
\begin{aligned}
G(\lambda) &\triangleq \int_{r(\lambda)}^{1} (1-\phi)\big(\mu_\lambda(z)X-1\big)+b \ \mathrm{d}z, \\
&= (1-r(\lambda))\left[(1-\phi)\left(\mu_\lambda\left(\tfrac{r(\lambda)+1}{2}\right)X-1\right)+b\right].
\end{aligned} \tag{92}
$$

The first derivative of $V_T^B(\pi)$ is :

$$
\begin{aligned}
\frac{\partial V_T^B}{\partial \pi} &= \widehat{\lambda}'(\gamma(\pi))\gamma'(\pi)\left[(1-\phi)\left(\tfrac{1}{2}X-1\right)+b-(1-\pi)\,G\big(\widehat{\lambda}(\gamma(\pi))\big)\right]-\int_{\widehat{\lambda}(\gamma(\pi))}^{1} G(\lambda)\,\mathrm{d}\lambda, \\
&= \widehat{\lambda}'(\gamma(\pi))\gamma'(\pi)\Big[1-(1-\pi)(1-\gamma(\pi))\Big]b-\int_{\widehat{\lambda}(\gamma(\pi))}^{1} G(\lambda)\,\mathrm{d}\lambda, \\
&= \frac{\overline{\lambda}}{(1-\sqrt{\pi})^2}\,b-\int_{\widehat{\lambda}(\gamma(\pi))}^{1} G(\lambda)\,\mathrm{d}\lambda,
\end{aligned}
$$

where the first line uses the Leibnitz integral rule, the second line follows from plugging $\widehat{\lambda}(\gamma(\pi))$ in Equation (92) and using Equation (42) and Equation (43), while the last line follows from the fact that

$$
\begin{aligned}
\widehat{\lambda}'(\gamma(\pi))\gamma'(\pi) &= \frac{\overline{\lambda}}{\sqrt{\pi}(1-\sqrt{\pi})^2} > 0, \\
(1-\pi)(1-\gamma(\pi)) &= 1-\sqrt{\pi}.
\end{aligned} \tag{93}
$$

Note that the left derivative of $V_T^B(\pi)$ at $\pi = \big(\frac{1}{X}/(1-\frac{1}{X})\big)^2$ is

$$
\frac{\partial V_T^B}{\partial \pi}\Bigg|_{\pi=\left(\frac{\frac{1}{X}}{1-\frac{1}{X}}\right)^2} = \frac{\overline{\lambda}}{\left(1-\frac{\frac{1}{X}}{1-\frac{1}{X}}\right)^2} b > 0, \tag{94}
$$

since $\widehat{\lambda}(\gamma(\pi)) = 1$ for $\pi = \big(\frac{1}{X}/(1-\frac{1}{X})\big)^2$.

The second derivative of $V_T^B(\pi)$ is

$$
\begin{aligned}
\frac{\partial^2 V_T^B}{\partial \pi^2} &= \frac{\overline{\lambda}}{\sqrt{\pi}(1-\sqrt{\pi})^3} b + \widehat{\lambda}'(\gamma(\pi))\gamma'(\pi)\,G\big(\widehat{\lambda}(\gamma(\pi))\big), \\
&= \frac{\overline{\lambda}}{\sqrt{\pi}(1-\sqrt{\pi})^2}\left[b\left(\tfrac{1}{1-\sqrt{\pi}}+1-\gamma(\pi)\right)+(1-\phi)\left(\tfrac{1}{2}X-1\right)\left(1+\frac{\sqrt{\pi}}{1-\sqrt{\pi}}\right)\right] > 0,
\end{aligned} \tag{95}
$$

where the first line follows from Leibnitz integral rule, while the second line follows from plugging $\widehat{\lambda}(\gamma(\pi))$ in Equation (92) and using Equation (42), (43) and Equation (93).

Step 3. From the shapes of $V_T^B(\pi)$ and $V_O^B(\pi)$ described above, it follows that:

- if $\phi \geq 1-2b$, we have
  - $V_O^B(\pi) > V_T^B(\pi)$ for $\pi < \big(\frac{1}{X}/(1-\frac{1}{X})\big)^2$,
  - $V_O^B(\pi) = V_T^B(\pi)$ for $\pi \geq \big(\frac{1}{X}/(1-\frac{1}{X})\big)^2$,

- if $\phi < 1 - 2b$ we have $V_O^B(\pi) > V_T^B(\pi)$ for every $\phi \in (0, 1)$.

■

Appendix B

This section provides closed form expressions of variables that can be observed by econometricians, such as the amount of data shared and overall credit provision.

Proposition A.1 (Observables)**.** *The amount of data shared in equilibrium with a transparent and an opaque algorithm are, respectively:*

$$\begin{aligned} S_T(\pi, X) &= (1-\pi)\left[\widetilde{\lambda}(\pi)\big(1-\gamma(\pi)\big) + \int_{\widetilde{\lambda}(\pi)}^1 \big(1 - r(\lambda)\big)\mathrm{d}\lambda\right], \\ S_O(\pi, X, b, \phi) &= (1-\pi)2\eta(b, \phi); \end{aligned} \tag{96}$$

while the levels of credit provision under transparency and opacity are, respectively:

$$\begin{aligned} I_T(\pi, X) &= \widetilde{\lambda}(\pi) + (1-\pi)\int_{\widetilde{\lambda}(\pi)}^1 \int_{r(\lambda)}^1 \mathrm{d}z\,\mathrm{d}\lambda, \\ I_O(\pi, X, b, \phi) &= \pi + (1-\pi)\left[\overline{\lambda} + \int_{\overline{\lambda}}^1 \int_{\min\{r(\lambda), \eta(b,\phi)\}}^1 \mathrm{d}z\,\mathrm{d}\lambda\right]. \end{aligned} \tag{97}$$

Proof of Proposition A.1. Consider a data sharing strategy of the form

$$m_Q(z) = \begin{cases} \varnothing & \text{if } \ z \in Q, \\ z & \text{if } \ z \notin Q, \end{cases}$$

where $Q \in \{\mathcal{G}(\lambda, \pi), \mathcal{H}(b, \phi)\}$ is the data-withholding set induced by a specific transparency regime, and where $\mathcal{G}(\lambda, \pi)$ and $\mathcal{H}(b, \phi)$ are defined in Proposition 1 and Proposition 2, respectively.

The ex-ante level of equilibrium data-sharing from a transparency regime $\tau \in \{T, O\}$ inducing a data-withholding set Q is:

$$\begin{aligned} S_\tau(\pi) &\triangleq \frac{1}{2}\int_{-1}^1 S_Q(\pi, \lambda)\,\mathrm{d}\lambda, \\ &= \int_0^1 S_Q(\pi, \lambda)\,\mathrm{d}\lambda, \end{aligned} \tag{98}$$

where

$$S_Q(\pi, \lambda) \triangleq (1-\pi)\int_0^1 \mathbb{1}\left\{m_Q(z) = z\right\}\mathrm{d}z, \tag{99}$$

where the second line of Equation (98) follows from $S_Q(\pi, \lambda)$ being even in λ. Proceeding as in Proposition 3, after simple algebra we obtain the expression in Equation (96).

The ex-ante equilibrium level of credit provision from a transparency regime $\tau \in \{T, O\}$ inducing a data-withholding set Q is:

$$I_\tau(\pi) \triangleq \frac{1}{2}\int_{-1}^{1} I_Q(\pi,\lambda)\,\mathrm{d}\lambda, \\ = \int_{0}^{1} I_Q(\pi,\lambda)\,\mathrm{d}\lambda, \tag{100}$$

where

$$I_Q(\pi,\lambda) \triangleq \pi\ell_\lambda(\varnothing) + (1-\pi)\int_0^1 \ell_\lambda\big(m_Q(z)\big)\mathrm{d}z, \tag{101}$$

where $\ell_\lambda(z)$ and $\ell_\lambda(\varnothing)$ are defined in Equation (6) and Equation (21), while the second line of Equation (100) follows from $I_Q(\pi,\lambda)$ being even in λ. Proceeding as in Proposition 3, after simple algebra we obtain the expression in Equation (97). ■