BANK OF ENGLAND

Staff Working Paper No. 930

The impact of machine learning and big data on credit markets

Peter Eccles, Paul Grout, Paolo Siciliani and
Anna (Ania) Zalewska

July 2021

BANK OF ENGLAND

Staff Working Paper No. 930
The impact of machine learning and big data on credit markets

Peter Eccles,[(1)] Paul Grout,[(2)] Paolo Siciliani[(3)] and Anna (Ania) Zalewska[(4)]

Abstract

There is evidence that machine learning (ML) can improve the screening of risky borrowers, but the empirical literature gives diverse answers as to the impact of ML on credit markets. We provide a model in which traditional banks compete with fintech (innovative) banks that screen borrowers using ML technology and show that the impact of the adoption of the ML technology on credit markets depends on the characteristics of the market (eg borrower mix, cost of innovation, the intensity of competition, precision of the innovative technology, etc.). We provide a series of scenarios. For example, we show that if implementing ML technology is relatively expensive and lower-risk borrowers are a significant proportion of all risky borrowers, then all risky borrowers will be worse off following the introduction of ML, even when the lower-risk borrowers can be separated perfectly from others. At the other extreme, we show that if costs of implementing ML are low and there are few lower-risk borrowers, then lower-risk borrowers gain from the introduction of ML, at the expense of higher-risk and safe borrowers. Implications for policy, including the potential for tension between micro and macroprudential policies, are explored.

Key words: Adverse selection, banking, big data, capital requirements, credit markets, fintech, machine learning and prudential regulation.

JEL classification: G21, G28, G32, G28, 031, 033.

(1) Bank of England. Email: peter.eccles@bankofengland.co.uk
(2) Bank of England and University of Bristol. Email: paul.grout@bankofengland.co.uk
(3) Bank of England. Email: paolo.siciliani@bankofengland.co.uk
(4) University of Bath. Email: A.Zalewska@bath.ac.uk

We would like to thank the participants of the Conference on 'Financial innovation: implications for competition, regulation and monetary policy' organised by the Bank of England, Bank for International Settlements and CEPR, participants of the Bank of England Seminar and of the ACPR Research Initiative Seminar organised by the Banque de France for useful comments and suggestions. In particular, we are very grateful to Bill Francis and Neeltje van Horen for their input. Ania Zalewska would like to acknowledge that this article is based upon work from COST Action 19130, supported by COST (European Co-operation in Science and Technology), www.cost.eu.

The Bank's working paper series can be found at www.bankofengland.co.uk/working-paper/staff-working-papers

Bank of England, Threadneedle Street, London, EC2R 8AH
Email enquiries@bankofengland.co.uk

ISSN 1749-9135 (on-line)

1 Introduction

It is well documented that smaller and riskier businesses struggle to obtain finance at reasonable terms relative to their larger and more stable counterparts (e.g. Tang (2009), Ayygai *et al.* (2010), Canales & Nanda (2012), Rodano *et al.* (2018)), but it is anticipated that the situation will improve significantly with banks' adoption of machine learning (ML) methodologies and availability of Big Data. Expectations of beneficial change are high, but many important questions arise, e.g. what is the likely impact on risk premia, which borrowers are targeted by lenders adopting ML and Big Data methodologies, will lower risk borrowers suffer, will risk increase in the banking sector? The empirical literature suggests that there is no unique answer to most of these questions at the present time (e.g. Jagtiani & Lemieux (2019), Dorfleitner *et al.* (2016)). Of course, it is not obvious that the answers should be unique since the outcomes will depend on the nature of the market and the innovation. This begs the question of whether there are simple intuitive relationships between the impact of ML and the characteristics of the market (e.g. the intensity of competition, borrowers' mix, cost of innovation, precision of the innovative technology, etc.). This paper provides a simple model that elucidates these relationships.

According to Gambacorta *et al.* (2019) in many countries the value of the domestic credit granted by domestic financial institutions significantly exceeds GDP.[1] Yet, regardless of country's development, small and medium size enterprises (SMEs) struggle to obtain the capital necessary to finance their growth. In the U.S. in 2015, 50% of small companies (less than 500 employees) and 58% of startups reported a financing shortfall.[2] In China, only 58.36% of loan amounts requested were finally granted.[3] In the UK, 30% of SMEs reported that they faced difficulties in accessing the funding they needed,[4] and according to the survey conducted by the British Business Bank, UK SMEs "seeking bank finance are more likely to say the reason for their worse than expected experience was due to high interest rates".[5]

Thus, given that in many countries credit markets are much bigger than corresponding

[1]For instance, in 2018, the value of the domestic credit granted by domestic financial institutions amounted to 173% of GDP (or $21.1 trillion) in the U.S.A., 296% ($13.9 trillion) in Japan, 219% (or $4.6 trillion) in the UK, 256% (or $30.2 trillion) in China (Gambacorta *et al.* (2019)).

[2]https://www.pymnts.com/news/b2b-payments/2016/smes-severely-dissatisfied-with-elending-experience-fed/

[3]https://www.oecd-ilibrary.org/sites/31f5c0a1-en/index.html?itemId=/content/component/31f5c0a1-en

[4]https://www.merchantsavvy.co.uk/uk-sme-data-stats-charts/

[5]https://www.british-business-bank.co.uk/wp-content/uploads/2020/03/2019-Business-Finance-Survey.pdf

economies and stock markets, and the importance of SMEs for employment, any inefficiencies in capital allocation can have a significant impact on economic and social development. The potential of ML algorithms to better identify risks and differences in risks between those seeking credit, may be a game changer that improves credit allocation (e.g. Padhi (2017), de Roure *et al.* (2019), Jagtiani & Lemieux (2019), Fu *et al.* (2020)) and lowers the cost of borrowing (e.g. Philippon (2016), Albanesi & Vamossy (2019), Jagtiani & Lemieux (2019), Fu *et al.* (2020)). Yet, there is a growing body of literature raising doubts about the benefits of ML adoption in credit markets (e.g. Milone (2019), Parlour *et al.* (2020)). However, when the effects of ML adoption are discussed, it is done either from the perspective of individual institutions who have innovated or as a comparison between 'traditional' banks that do not innovate and non-banking competitors such as P2P (peer to peer) platforms. To the best of our knowledge this is the first paper that provides a model in which both 'traditional' banks and innovators are subject to the same regulatory requirements and risk-taking restrictions, i.e. the innovators are also banks.

In our model, there are three types of borrowers: safe borrowers, borrowers with high repayment risk and borrowers with low repayment risk. There also are two types of banks, traditional and innovative. Traditional banks can separate safe from risky borrowers but cannot distinguish between low repayment risk and high repayment risk borrowers. Innovative banks, i.e. those using Big Data and ML methodology, are much better at screening potential borrowers than the traditional banks. That is, innovative banks, in addition to being able to separate safe from risky borrowers, are also good at separating low repayment risk from high repayment risk borrowers. Adopting Big Data and ML methodologies, which we often simply refer to as 'innovation', is not cost free. If costs of innovation are high for most banks, hence the innovative sector is small relative to traditional banks, innovative banks are, at best, only able to 'cream off' a small proportion of the borrowers with low repayment risk. In this case, the dominant form of competition in the market remains that between the traditional banks. By pulling a disproportionate number of low risk projects away from traditional banks, innovative banks create an adverse selection problem for traditional banks, which creates a negative externality for all risky borrowers.[6] This raises repayment rates for all risky borrowers, chokes off demand for risky loans and results in a 'flight to safety'. In this situation, the gains to innovation can be high but risky borrowers of both types are worse off and it is the safe borrowers who benefit from a decline in repayment rates. In contrast, if costs of innovation are low (hence the innovative sector is much larger) or if most of the

[6] Vallée & Zeng (2019) identify a similar adverse selection effect in the case of P2P platforms.

risky borrowers in the market have high repayment risk, then competition between innovative banks bites and is the central driver of repayment rates. The competition among innovative banks pushes down the repayment rates for low repayment risk borrowers and sucks more low repayment risk borrowers into the market. In this situation, innovative banks bring significant benefits to the lower repayment risk borrowers. The proportion of lower repayment risk borrowers in the market is increased and innovative banks are more complementary to, rather than competing directly against, traditional banks. However, the gains from innovation for innovative banks are lower than in the previously described case (i.e. when the innovative sector remained small).

The above examples have assumed that innovative banks opt for 'better' projects (cream skimming), and this is the case whenever the ML technology is good at identifying different risk types. However, we show that if the technology is not very accurate at identifying risk types of potential borrowers, then the innovative banks can prefer to lend to the higher repayment risk borrowers (bottom fishing) rather than the lower repayment risk borrowers (cream skimming). Among other results we also show that (network externality) multiple equilibria can arise. In addition, we consider special cases where the costs for innovators are the result of cost of access to Big Data.

This research adds to several strands of the literature. First, it contributes to the literature documenting which lenders benefit from the presence of innovative/Fintech lenders and the closely related problem of whether these lenders target high risk or low risk borrowers. Papers find different outcomes, some finding that the benefit is mostly concentrated on more risky borrowers (e.g. de Roure *et al.* (2019), Balyuk *et al.* (2020)), some claiming that there is no evidence that innovative lenders have targeted higher risk borrowers (e.g. Fuster *et al.* (2020)) or that there is evidence of cream skimming (e.g. Padhi (2017)). We show that there is no standard answer as to who may gain. Our model illustrates that, according to the cost of innovation and the mix of high and low risk projects in the population, the beneficiaries of the presence of innovative banks can be any of the three types of borrower we consider. We identify how the different outcomes depend on the size of the innovative sector (which depends on the cost of innovation), the mix of borrower types in the market and the quality of the innovative technology.

The model also speaks to the literature on the financial benefits of innovation. Some papers document significant benefits to Fintech (e.g. Chen *et al.* (2019)) whilst others suggest it is too soon to tell (Freedman & Jin (2017), Claessens *et al.* (2018)). Others have identified the lack of low-cost finance to the sector as a limitation on its growth, suggesting the financial gains of being in the sector may be limited (e.g. (FSB) (2019). We identify factors that influence the returns to adopting ML screening of borrowers and show that the benefits may be related to the dominant competitive forces in the market and negatively correlated with the gains that the sector brings to borrowers.

We also contribute to the literature that examines whether the presence of innovative lenders is harmful rather than beneficial for borrowers and for society. Papers that are concerned with the negative impact of ML are often focused on discrimination (e.g. Fu *et al.* (2020), Fuster *et al.* (2020)). There are many papers that emphasize the positive benefits of ML (e.g. Lessmann *et al.* (2015), Iyer *et al.* (2016)) and others that question whether the technological effect is significant (Hand (2006)). The model shows which characteristics of the market determine whether the net effects are positive or negative.

Finally, we add to the literature that discusses the expansion and structural change of the lending market. Much of the evidence focuses on P2P lending. It is common in the literature to conclude that the expansion of the market is channeled through non-banking institutions that do not face exogenously imposed regulatory costs (e.g. de Roure *et al.* (2019)). In our model, we compare innovative banks and traditional banks. Hence, both face the same regulatory framework and both are intermediaries. We show that even when the structure of the lenders is similar, so that the sole difference is in their innovation capability, we obtain a rich set of outcomes which depend on the market characteristics, i.e. they are not dependent on any differences in regulatory and intermediary status of the two types of lenders.

2 Literature review

It is often claimed that banking services are not available or are extremely expensive exactly where they are most needed. For instance, it is well-recognized that SMEs find it much harder than big corporations to access the capital required for their development (e.g. Kersten *et al.* (2017), Tang (2009), Ayygai *et al.* (2010), Behr *et al.* (2012), Canales & Nanda (2012), Rodano *et al.* (2018)), and that this is a particularly acute problem in countries with weak legal systems (e.g. Hazelmann & Wachtel (2010)). Moreover, evidence suggests that access to capital displays racial and gender discrimination (e.g. Fay & Williams (1993), Bellucci *et al.* (2010), Ongena & Popov (2016), Palia (2016), Galli *et al.* (2020)). Such inequality and unfairness in access to capital can have strong, negative economic consequences (e.g. Berg (2018), Frost *et al.* (2019), Beck *et al.* (2020)).

The limited quantity, and frequently poor quality, of credit history information of small borrowers is often blamed for the difficulties they face in getting loans. Goel *et al.* (2021) argue that, all things being equal, banks that base their lending decisions on 'hard' information grow faster and bigger than banks that make 'soft' information the core of their business model. Consequently, lending based on hard information is preferred if the information is available. If big banks have enough business with hard-information borrowers, they will have little incentive to develop lending services directed towards clients such as start-ups, small businesses and individuals who cannot provide sufficient, good quality hard information. Small banks are, therefore, 'pushed' towards such borrowers, and are more likely to use soft information in their lending decisions. Given that relational banking based on soft information is more expensive, it is also associated with higher repayment rates (Goel *et al.* (2021)). On a positive note, relational banks provide continuous funding at preferential rates in a crisis (Bolton *et al.* (2016), Karolyi (2018)) and work well for borrowers in close proximity to the lender (Granja *et al.* (2018)).

However, as the digital footprints left by individuals becomes richer, can the adoption of ML methodologies and Big Data help improve the screening of potential borrowers? If the adoption of ML methodologies can deliver more accurate assessments of risk profiles, will it improve the allocation of capital? Who will the main beneficiaries be? If it is borrowers who gain, which borrowers will benefit most? Some of these questions are being debated in the literature, but the research seems to deliver answers that often contradict each other.

The benefit of adopting ML algorithms to improve the quality of assessment of the riskiness of borrowers is broadly documented. Berg *et al.* (2020) claim that basic information left by individuals simply by accessing or registering with websites is rich enough to deliver projections of default rates consistent with credit bureau scores. Moreover, the adoption of ML algorithms enables more accurate assessment of risk profiles of potential borrowers (e.g. Iyer *et al.* (2016), Albanesi & Vamossy (2019), Jagtiani & Lemieux (2019), Parlour *et al.* (2020)), opens credit market to borrowers who would remain excluded if ML innovation was not adopted (e.g. Padhi (2017), de Roure *et al.* (2019), Jagtiani & Lemieux (2019), Fu *et al.* (2020)), lowers the cost of borrowing (e.g. Philippon (2016), Albanesi & Vamossy (2019), Jagtiani & Lemieux (2019), Fu *et al.* (2020)), and benefits the lenders that adopt ML techniques (e.g. Albanesi & Vamossy (2019), Chen *et al.* (2019), Fu *et al.* (2020)).

There also are papers, however, that show a less optimistic side of the adoption of the ML innovation. For instance, Fernández-Delgado *et al.* (2014) document that there are substantial differences in the quality of ML models and that "a few models are clearly better than the remaining ones". Hand (2006) argues that ML models often fail to take into account important aspects of real problems, so that the apparent superiority of more sophisticated methods may be something of an illusion. However, problems with the adoption of ML methodologies are not limited to programming aspects. There is growing evidence that using ML techniques does not benefit all groups of borrowers in the same way. For instance, racial discrimination reported in the traditional banking literature is also documented for ML-enhanced lending. Fuster *et al.* (2020) argue that black and hispanic borrowers are disproportionately less likely to gain from the introduction of ML, and that relying on ML algorithms increases disparity in rates between and within ethnic groups. Padhi (2017) finds evidence that ML-enhanced lending increased approval rates for black borrowers, yet, lowered loan ratings for approved borrowers who lived in areas with more black residents. Fu *et al.* (2020) confirmed racial and gender bias of ML algorithms even though race and gender were not used as algorithm inputs.

Indeed, it seems that even though the adoption of ML algorithms benefits some borrowers, the benefits are not universal. There is considerable evidence that the adoption of screening technologies based on ML algorithms leads to better processing of the existing hard information rather than utilizing soft information (Balyuk *et al.* (2020), Di Maggio & Yao (2020)), but Dorfleitner *et al.* (2016) show that, although hard information is preferred by Fintech lenders, they can effectively utilize the existing soft information. Berg *et al.* (2020) argue that ML innovation improved access to funding for borrowers with a good digital footprint but reduced it for those with poor digital footprints. Milone (2019) concludes that using information-based screening methods reduces incentives to collect soft information by technologically advanced lenders. This disadvantages borrowers with limited historical data who have no choice but to rely on traditional lenders and pay higher borrowing rates. Vallée & Zeng (2019) show that, when sophisticated lenders screen the market, repayment rates increase.[7] In contrast, Parlour *et al.* (2020) claim that the adoption of Fintech technologies reduces the repayment rates for

[7]Vallée & Zeng (2019) do not discuss the consequences of adopting financial innovation *per se*, rather they discuss sophisticated lenders who screen borrowers and lenders who do not. However, the sophisticated lenders can be perceived as lenders who are capable of screening borrowers (e.g. using ML algorithms), and unsophisticated lenders who are without such capabilities.

those borrowers who did not have strong connections (affinity) with banks but increases the repayment rates for borrowers with well-established relations. de Roure *et al.* (2019) come to a somewhat similar conclusion, i.e. they claim that the adoption of Fintech technologies reduces the borrowing rates for more risky borrowers. Di Maggio & Yao (2020) find evidence that the reduction of borrowing rates for Fintech's clients is short-lived, as the Fintech's clients have higher default rates and subsequently Fintech lenders price it in their products.

Given the wide variety of conclusions, it is not clear whether Fintech lenders better screen and cream-skim the less risky borrowers or whether, deliberately or by 'programming fault', they facilitate the financing of more risky borrowers. Padhi (2017) concludes that Fintech (P2P) lenders 'cherry pick' the best loans, making the portfolios of bank loans riskier. Also, Parlour *et al.* (2020) claim that the existence of Fintech non-banking competitors increases risk taking of the traditional banks. In contrast, de Roure *et al.* (2019) conclude that the Fintech lenders target borrowers who would not succeed in securing a loan from 'traditional' banks since these only specialize in lending to safe borrowers. Also, Dorfleitner *et al.* (2016), Di Maggio & Yao (2020) and Fu *et al.* (2020) find evidence that non-banking intermediaries that adopt Fintech increase risk.

The above papers discuss the impact of the adoption of ML techniques when the innovation is introduced to the market by non-banking intermediaries. These non-banking intermediaries benefit from having more flexibility when it comes to risk taking, since they do not have to comply with capital requirements and other regulation to restrict risk-taking. Whilst P2P lending, crowdfunding, and other forms of non-bank financing play an extensive role in some economies, they are still very small in relation to the size of the banking sector and scale of lending provided by banks, and Fintech innovation is not limited to non-banking lenders. Many banks have adopted ML-based screening of borrowers, and non-banking lenders are choosing to become banks (e.g. Zopa in the UK). Therefore, it is important to consider how borrowers are affected when some banks adopt ML methodologies and, hence, become better at assessing risk of potential borrowers, whereas other banks follow their traditional business model.

In a traditional bank setting, increasing competition reduces the probability of individual borrowers' default (e.g. Boyd & De Nicolo (2005)). However, it is also likely to reduce the profit to a bank on those projects that succeed, and thus reduce the cushion from successful projects to cover the losses from failing projects. The net effect on bank stability depends on the balance between these two forces (see Martinez-Miera & Repullo (2010)). Both Boyd & De Nicolo (2005) and Martinez-Miera & Repullo (2010) have competition for borrowers from symmetrically positioned banks. Boot & Thakor (2002) argued that an increase in competition among banks would encourage banks to shift from transaction to relationship lending, however, this does not seem to be the case when the competition is driven by differences in adoption of new technologies. In this paper, we analyze the impact of asymmetric forms of competition. The impact of competition from innovative banks on the stability of the banking system depends on the specific circumstances, but the model gives insight as to which factors matter and when.

3 Model

We assume that there are many banks who can lend to three types of projects: safe projects S, and two types of risky projects, type A and type B, and that *ex ante* borrowers do not know if their project is type A or type B. All projects of a given type are identical. Safe projects never fail, i.e. the borrowers can always repay the money borrowed from banks. However, type A and type B projects have non-zero probabilities of failure, p_A and p_B respectively, and the probability that both of them fail equals p_{AB}. When type A projects fail, they repay either $r_{A,F}$ or the repayment rate (whichever is lower). Analogously, when type B projects fail they repay $r_{B,F}$ or the repayment rate (whichever is lower). Both type A and type B projects repay the full repayment rate when they succeed.

We assume that (i) $r_{A,F} \geq r_{B,F}$, (ii) $p_A \leq p_B$ and (iii) at least one of these two inequalities is strict, i.e. from the banks' perspective, type A projects have a more attractive profile than type B projects.

Innovative banks hold a proportion $\mu \in [0, 1]$ of the total funds held by banks, with the remaining funds (i.e $1 - \mu$) being held by traditional banks. Traditional banks can identify safe projects but cannot distinguish between type A projects and type B projects. Innovative banks can distinguish between all project types. Each bank chooses a set of repayment rates. In particular, banks offer repayment rates r_θ^j, where $j \in \{I, T\}$ indicates whether the rate is offered by an innovative bank (I), or a traditional bank (T), and $\theta \in \{S, A, B\}$ indicates the type of project. Since traditional banks cannot distinguish between the two types of risky project, it must be the case that $r_A^T = r_B^T$ and we use r_R^T to refer to this repayment rate.

We also assume that, in aggregate, banks have a total quantity of funds D to allocate to the projects. For each bank, a proportion of funds q must be provided by capital investors while remaining funds are provided by depositors. Depositors are always paid a deposit rate s and deposit rate payments are funded primarily from the revenue of the relevant bank. If a bank's revenue from projects is insufficient to cover its deposit payment obligations, the difference between bank revenues and promised payments to depositors is met by a government funded deposit insurance scheme. To provide a potential role for deposit insurance, we assume $r_{B,F} < s$.

The aggregate demand for loans $\Phi_S(.)$, $\Phi_A(.)$ and $\Phi_B(.)$ from projects of type S, A and B respectively, depends on the competitive repayment rates they face. We assume that all the demand functions Φ_S, Φ_A, Φ_B are strictly decreasing and continuous, i.e. the borrowers are less likely to borrow funds when they are required to pay a higher repayment rate. The proportion of type B projects among the risky projects in the portfolios of the traditional banks is denoted by ζ, and hence a higher value of ζ means that the risky projects financed by traditional banks have a larger proportion of type B projects. Finally, note that repayment rates and mix of risky project parameter ζ may vary as the proportion of funds held by innovative banks μ changes. We use $r_\theta|_{\mu=\mu_c}$ to denote the repayment rate offered to projects of type θ given $\mu = \mu_c$ and similarly use $\zeta|_{\mu=\mu_c}$ to denote the value of the mix of risky project parameter given $\mu = \mu_c$.

4 Analysis

In this section we consider three scenarios which capture different degrees of innovation in the banking sector. The first scenario we call the 'increasing risk premia' scenario. In this scenario there is a risk premium for all risky projects with the repayment rate for risky projects increasing and for safe projects decreasing as the size of the innovative sector increases. We then have an 'intermediate scenario'. In this scenario, there also is a single repayment rate for all risky projects but the repayment rate for risky projects is weakly decreasing as the size of the innovative sector increases. The repayment rate for safe projects is weakly increasing. The final scenario is called a 'segregated scenario'. In this scenario traditional banks do not hold type A projects and there are separate repayment rates for each of the project types. Initially, in this scenario, the repayment rates for type A projects decline as the size of the innovative sector increases, whilst the repayment rates for safe and type B projects increase.

Eventually, if the level of innovation is high, then there are no further gains from innovation in the banking sector and the repayment rates are constant as the size of the innovative sector increases.

The thresholds between the scenarios are denoted:

- μ^* - the boundary between the increasing risk premia scenario and the intermediate scenario,
- μ^{**} - the boundary between the intermediate scenario and the segregated scenatio.

This section first describes the characteristics of the scenarios for each μ (exogenously given, assuming there are no costs of becoming an innovative bank). We then include a cost of innovation, which endogenises the proportion of innovative banks. However, before we do either, it is useful, as a benchmark, to address the situation where there are no innovative banks.

4.1 No innovative banks

This is the situation when $\mu = 0$, i.e. there are only traditional banks and by assumption these cannot separate type A projects from type B projects. Thus, we refer only to safe projects S and risky projects R. In this case, in equilibrium, the repayment rates the traditional banks charge borrowers with safe projects are r_S^T, and the repayment rates of type A and type B projects are the same, and denoted r_R^T. We first define $\mathcal{L}(r_R^T, \zeta)$ to be the expected loss suffered by a bank investing in only risky projects (when the repayment rate paid by the risky projects equals r_R^T and a proportion ζ of these projects are type B projects). In particular:

$$
\begin{aligned}
\mathcal{L}(r_R^T, \zeta) \quad = \quad & (p_A - p_{AB})\Big(r_R^T - \max\{(1-\zeta)r_{A,F} + \zeta r_R^T, (1-q)s\}\Big) \\
& +(p_B - p_{AB})\Big(r_R^T - \max\{(1-\zeta)r_R^T + \zeta r_{B,F}, (1-q)s\}\Big) \\
& +p_{AB}\Big(r_R^T - \max\{(1-\zeta)r_{A,F} + \zeta r_{B,F}, (1-q)s\}\Big). \qquad (1)
\end{aligned}
$$

The expected loss $\mathcal{L}(r_R^T, \zeta)$ suffered by a bank investing in risky projects depends on the deposit insurance. If in a given state of the world a project cannot repay the agreed repayment rate r_R^T, the project will repay what it can. If this is above what is required to repay depositors, then the bank loses the difference between the required repayment and what it receives. However, if there are insufficient funds to repay depositors, then the maximum the bank can lose is the difference between what projects should have repaid (i.e. r_R^T) and the amount that depositors receive (i.e. $(1-q)s$). Any shortfall beyond this is picked up by the deposit insurance. Repayment rates are set in equilibrium such that the expected return on the safe projects S is equal to the expected return from the risky projects. In particular, this means that the revenue paid by the safe projects r_S^T equals the maximum revenue repaid by the risky projects r_R^T less expected losses $\mathcal{L}(r_R^T, \zeta)$. It follows that:

$$r_S^T = r_R^T - \mathcal{L}(r_R^T, \zeta). \tag{2}$$

This shows that $\mathcal{L}(r_R^T, \zeta)$ is the risk premium traditional banks demand to invest in risky projects. In equilibrium all financial resources of the banks are allocated, i.e. the market clears and hence:

$$\Phi_S(r_S^T) + \Phi_A(r_R^T) + \Phi_B(r_R^T) = D. \tag{3}$$

Finally, the share of the risky projects of type B, ζ, equals:

$$\zeta|_{\mu=0} = \frac{\Phi_B^T(r_R^T|_{\mu=0})}{\Phi_A^T(r_R^T|_{\mu=0}) + \Phi_B^T(r_R^T|_{\mu=0})}. \tag{4}$$

Hence, when there are no innovative banks the repayment rates charged to safe and risky projects (r_S^T and r_R^T, respectively) and the mix of risky projects, ζ, satisfy equations (2), (3) and (4).

Consider a situation where a bank invests only in risky projects, all of which face repayment rate of $r_R|_{\mu=0}$. We define γ_B as the maximum proportion of type B projects that such a bank can hold while still ensuring that it will remain solvent when (i) type A projects succeed and (ii) type B projects fail. Given that we assume $r_{B,F} < (1-q)s$, the threshold γ_B satisfies:

$$(1-\gamma_B) r_R^T|_{\mu=0} + \gamma_B r_{B,F} = (1-q)s. \tag{5}$$

Rearranging Eq. (5) gives:

$$\gamma_B = \frac{r_R^T|_{\mu=0} - (1-q)s}{r_R^T|_{\mu=0} - r_{B,F}}. \tag{6}$$

We now discuss the scenario where some banks are innovative, i.e. $\mu > 0$.

4.2 The increasing risk premia scenario

First, we consider the scenario in which $0 < \mu \leq \mu^*$, for some μ^*. In this case, the proportion of innovative banks is low relative to the proportion of type A projects available in the market. For this reason, the repayment rates in the marketplace will be driven by competition between traditional banks. Traditional banks cannot distinguish between type A and type B projects and, therefore, in equilibrium, traditional banks offer all risky projects that borrow from traditional banks the same repayment rate r_R^T. The increasing risk premia scenario has the following relationship between the repayment rates:[8]

Proposition 4.1 *Suppose $\zeta|_{\mu=0} < \gamma_B$, i.e. traditional banks are solvent when (i) there are no innovative banks (i.e. $\mu = 0$) and (ii) the type A projects succeed. In this case there exists $\mu^* > 0$ such that whenever $0 \leq \mu < \mu^*$:*

- *ζ strictly increases as μ increases,*

- *r_S^T strictly decreases as μ increases,*

- *$r_R^T = r_R^I$ strictly increases as μ increases.*

To see why Proposition 4.1 holds, first note that innovative banks are only incentivised to offer loans to type A projects. Given that,

$$\zeta = \frac{\Phi_B^T(r_R^T)}{\Phi_B^T(r_R^T) + \Phi_A^T(r_R^T) - \mu D}, \tag{7}$$

the proportion of type B projects in the pool of projects financed by the traditional banks increases as μ increases.

To explain the results concerning the change in repayment rates, note that traditional banks are indifferent between investing in safe projects and the mix of risky projects available to them.

If μ is sufficiently small, then traditional banks remain solvent whenever type A projects succeed. Given this, we must be in one of the following three cases:

Case 1: Traditional banks remain solvent even if both type A and type B projects fail simultaneously. In this case the banks are able to repay depositors in all states of the world and hence, the risk premium takes the form:

[8] We assume a simple Riley equilibrium throughout (Riley (1979)).

$$\mathcal{L}(r_R^T,\zeta) = p_A(1-\zeta)(r_R^T - r_{A,F}) + p_B\zeta(r_R^T - r_{B,F}).$$

Differentiating it with regard to ζ while keeping r_R^T constant gives

$$\frac{\partial\mathcal{L}}{\partial\zeta} = p_B(r_R^T - r_{B,F}) - p_A(r_R^T - r_{A,F}) > 0.$$

Case 2: Traditional banks only become insolvent if both projects fail. In this case the risk premium takes the form:

$$\mathcal{L}(r_R^T,\zeta) = (p_A - p_{AB})(1-\zeta)(r_R^T - r_{A,F}) + (p_B - p_{AB})\zeta(r_R^T - r_{A,F}) + p_{AB}\Big(r_R^T - (1-q)s\Big).$$

Differentiating it with regard to ζ while keeping r_R^T constant gives

$$\frac{\partial\mathcal{L}}{\partial\zeta} = (p_B - p_{AB})(r_R^T - r_{B,F}) - (p_A - p_{AB})(r_R^T - r_{A,F}) > 0.$$

Case 3: Traditional banks become insolvent whenever type A projects fail. In this case the risk premium takes the form:

$$\mathcal{L}(r_R^T,\zeta) = (p_B - p_{AB})\Big(\zeta(r_R^T - r_{B,F})\Big) + p_A\Big(r_R^T - (1-q)s\Big).$$

Differentiating it with regard to ζ while keeping r_R^T constant gives

$$\frac{\partial\mathcal{L}}{\partial\zeta} = (p_B - p_{AB})(r_R^T - r_{B,F}) > 0.$$

The risk premium increases in ζ in each case because lending to a higher proportion of type B projects makes the traditional banks less profitable (holding the repayment rate constant). We have noted above that an increase in μ leads to an increase in ζ, and hence (assuming μ is sufficiently small) also leads to an increase in the risk premium $\mathcal{L}(r_R^T,\zeta)$, i.e. $\frac{\partial\mathcal{L}}{\partial\zeta} > 0$. In order for the traditional banks to continue to accept this increased risk, the difference between the safe repayment rate and the risky repayment rate must increase.

Note, that if both repayment rates increased, then the aggregate number of borrowers with risky and safe projects would decline contradicting the market clearing condition (Eq. (3)).

Similarly, if both rates decreased, then there would be an increase in the number of borrowers with safe and risky project which, again, contradicts the market clearing condition. Hence, the rate that traditional banks charge the borrowers with the safe projects (r_S^T) falls and the rate charged to the borrowers of the risky projects (r_R^T) rises. Innovative banks charge the highest repayment rate they can subject to finding sufficient number of borrowers with type A projects. If μ is less than the proportion of type A projects, then this rate is equal to the rate traditional banks charge their borrowers with risky projects.

As the size of the innovative sector increases, the proportion of type A projects in traditional banks' portfolio of risky projects declines. Since $r_{B,F}$ is less than $(1-q)s$, then there must come a point when the proportion of type A projects is insufficient to offset the type B projects' shortfall between $(1-q)s$ and $r_{B,F}$ when they fail. At this point, traditional banks that specialize in risky projects will be unable to repay their depositors whenever type B projects fail. If $r_{B,F}$ is strictly less than $(1-q)s$, we define μ^* as the minimum μ where this happens, i.e.

$$\mu^* = \arg\max_{\hat{\mu}\in[0,1]}\{(1-\zeta|_{\mu=\hat{\mu}})r_R^T|_{\mu=\hat{\mu}} + \zeta|_{\mu=\hat{\mu}} r_{B,F} \geq (1-q)s.\}$$

This captures the fact that at the critical threshold μ^* traditional banks investing only in risky projects remain solvent when type A projects succeed and type B projects fail.

The above discussion shows that subject to the assumption that $\zeta|_{\mu=0} < \gamma_B$ (i.e. when there are only traditional banks, these banks are able to repay depositors whenever type A projects succeeds), the introduction of innovative banks raises the risk premium to all risky projects at both the traditional and the innovative banks, and reduces the repayment rates for the safe projects. Note, if $\zeta|_{\mu=0} \geq \gamma_B$, then there is no increasing risk premium and we immediately enter the intermediate scenario which will be discussed next. Given that γ_B is increasing in q (see Eq. 6), this also suggests that the higher the capital requirements are, then the more likely there will be an increasing risk premia scenario.

4.3 The intermediate scenario

Assume $\zeta(\mu) > \zeta(\mu^*)$ and $r_{A,F} \geq (1-q)s)$. In this case, traditional banks investing only in risky projects cannot repay depositors when type A projects succeed and type B projects fail. The risk premium $\mathcal{L}(r_R^T, \zeta)$ equals:

$$\mathcal{L}(r_R^T, \zeta) = (p_A - p_{AB})(1-\zeta)(r_R^T - r_{A,F}) + p_B\Big(r_R^T - (1-q)s\Big).$$

The first term captures the losses incurred when only type A projects fail (and the banks do not become insolvent), while the second term captures the losses incurred when type B projects fail (taking into account that the banks become insolvent and shortfall in repayments needed to repay depositors is paid by the deposit insurance scheme). Note, that in this case the risk premium $\mathcal{L}(r_R^T, \zeta)$ decreases as ζ increases (holding r_R^T constant). This is because (i) avoiding losses associated with type A projects increases the profitability of traditional banks and (ii) increasing losses associated with type B projects does not affect bank profitability since when type B projects fail traditional banks always become insolvent. It follows that a small increase in the share of the innovative sector μ reduces the risk premium demanded by the traditional banks and hence reduce the repayment rates charged to the risky projects. This motivates the following result:

Proposition 4.2 *Suppose $r_{A,F} \geq (1-q)s > r_{B,F}$ (i.e. bank insolvency is possible only when type B projects fail). Then, there exists μ^{**} such that whenever $\mu^* \leq \mu < \mu^{**}$:*

- r_S^T strictly increases as μ increases,

- r_R^T strictly decreases as μ increases.

We now consider the case when $r_{A,F} < (1-q)s$. Now, there may exist $\mu_{AB}^* \in (\mu^*, \mu^{**})$ such that whenever $\mu \in (\mu^*, \mu_{AB}^*)$ traditional banks only remain solvent if both types of projects succeed. If this is the case, the result in Proposition 4.2 only holds over the interval $[\mu_{AB}^*, \mu^*]$ where traditional banks remain solvent when type A projects fail and type B projects succeed. Over the interval $[\mu^*, \mu_{AB}^{**}]$ traditional banks fail whenever either type A projects fail or type B projects fail, and in this case the risk premium becomes:

$$\mathcal{L}(r_R^T, \zeta) = (p_A + p_B - p_{AB})\Big(r_R^T - (1-q)s\Big).$$

In this case, the risk premium term does not depend on ζ. Since the risk premium term is constant in this interval, it follows that small increases in μ in the interval $[\mu^*, \mu_{AB}^*]$ do not change the repayment rates offered to the safe and the risky borrowers.

Taking the results in this subsection together, we observe the slightly counter-intuitive outcome that the higher $r_{A,F}$ is, then the larger is the interval where r_R^T is declining in μ. To put it another way, in the intermediate scenario, the greater the expected repayment by type A projects to traditional banks is, then the more likely it is that traditional banks will gain from having a smaller rather than a larger proportion of type A projects in their portfolio of risky projects.

4.4 The segregated scenario

If $\mu \geq \mu^{**}$, the innovative sector is big enough to absorb all borrowers with type A projects at the prevailing repayment rates (in particular, when $\Phi_A(r_R^T) = \mu D$). As μ increases above μ^{**}, innovative banks compete to lend to the borrowers with type A projects and offer them r_R^I repayment rate. The competition reduces the repayment rates offered by the innovative banks until enough borrowers with type A projects are brought into the market to meet the demand by innovative banks. For $\mu > \mu^{**}$ traditional banks know that their borrowers with risky projects are all type B and set the repayment rates to reflect this.

Proposition 4.3 *There exists μ_m such that whenever $\mu \in (\mu^{**}, \mu_m]$:*

- r_S^T *strictly increases with μ,*

- r_R^I *strictly decreases with μ,*

- r_R^T *strictly increases with μ.*

Whenever $\mu \in (\mu^{**}, \mu_m)$ borrowers with the safe projects borrow at the repayment rate r_S^T, borrowers with type A projects borrow at the repayment rate r_R^I, and borrowers with type B projects borrow at the repayment rate r_R^T. The overall market clearing condition is given by:

$$\Phi_S(r_S^T) + \Phi_A(r_R^I) + \Phi_B(r_R^T) = D,$$

and the market clearing condition for innovative banks equals:

$$\Phi_A(r_R^I) = \mu D.$$

This means that when the share of the innovative sector μ grows, i.e. the right hand-side of the above equality increases, the demand for borrowers with type A projects grows (i.e. the left-hand side of the equality increases). Since the demand function $\Phi_A(.)$ is a decreasing function of the repayment rates, it follows that r_R^I decreases as μ increases.

Consider now the market clearing condition for traditional banks:

$$\Phi_S(r_S^T) + \Phi_B(r_R^T) = (1-\mu)D.$$

It follows that when the level of innovation μ increases, then $\Phi_S(r_S^T) + \Phi_B(r_R^T)$ decreases. Because at this stage traditional banks only finance type B projects

$$r_S^T = r_R^T - p_B(r_R^T - (1-q)s).$$

It follows that r_S^T and r_R^T move in the same direction as μ increases. However, if both r_S^T and r_R^T decline or remain the same, then $\Phi_S(r_S^T) + \Phi_B(r_R^T)$ would not decrease (which as argued above is not the case). Given that $\Phi_S(r_S^T) + \Phi_B(r_R^T)$ decreases, then it must be true that as the share of the innovative sector increases, both r_S^T and r_R^T increase. In particular, as μ increases, the competition among innovative banks decreases r_R^I and increases the level of r_S^T until:

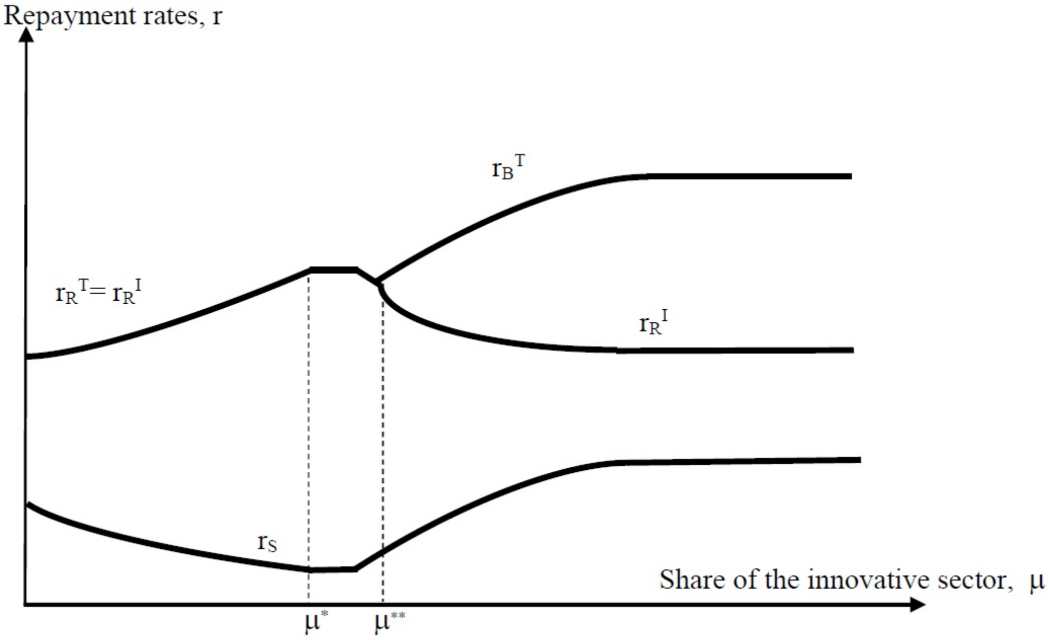
$$r_S^T = r_R^I - p_A(r_R^I - (1-q)s).$$

At this point, innovative banks are indifferent between investing in type A projects and the safe projects. Moreover, it is not optimal for innovative banks to lower the repayment rates any further. Thus, when $\mu > \mu_m$, the innovative sector has no incentive to grow any further and repayment rates are constant.

Putting all these propositions together gives a set of repayment rates as shown in Figure 1:

Note that Proposition 4.1 is contingent on the assumption that when there are no innovative banks (i.e. $\mu = 0$), traditional banks can repay depositors when type A projects succeed and type B projects fail. In reference to Figure 1, we should point out that if the proportion of type B projects is so large that with $\mu = 0$ traditional banks fail whenever type B projects fail, then there is no increasing risk premia scenario, i.e. $\mu^* = 0$.

Figure 1: Repayment rates of risky and safe projects as a function of the market share of the innovative sector μ



4.5 Cost to invest in data science capability

In this section we turn to the cost of innovation and assume that it costs bank j a fixed amount $c(j)$ to innovate. By relabeling banks if necessary, we ensure $c(j_1) \leq c(j_2)$ whenever $j_1 \leq j_2$. We define Δ to be the difference in expected profit between innovative banks and traditional banks, i.e.

$$\Delta = (1 - p_A) r_A^I + p_A \max\{(1-q)s, r_{A,F}\} - r_S^T. \tag{8}$$

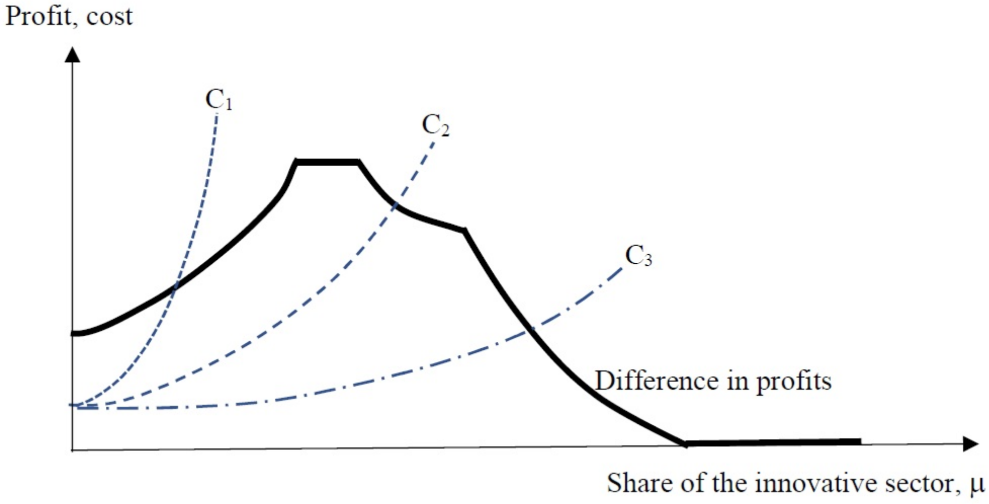
The value of Δ varies depending on the level of innovation μ. Let $\Delta|_{\mu=\mu'}$ denote the difference in profitability between innovative and traditional banks given the level of innovation $\mu = \mu'$. Note that $\Delta|_{\mu=0} < \Delta|_{\mu=\mu^*}$ because during the increasing risk premia scenario the difference between the repayment rates paid by the safe and the risky projects grows as the level of innovation increases. This in turn increases the difference in profitability between innovative banks investing in type A projects compared with traditional banks investing in safe projects.

We now state a result:

Proposition 4.4 *For every level of innovation $\mu' < \mu_m$, there exists a function $c(.)$ such that there is a unique equilibrium where the level of innovation equals μ'.*

To demonstrate this, first define $\Delta_L = \min\{\Delta|_{\mu=0}, \Delta|_{\mu=\mu'}\}$ and $\Delta_H = \Delta|_{\mu=\mu^*}$. Now consider a cost function $c(j)$ where (i) $c(j) < \Delta_L$ whenever $j \leq \mu'$ and (ii) $c(j) > \Delta_H$ whenever $j > \mu'$. We now consider banks with high and low costs of innovation in turn.

Figure 2: Difference in profits between an innovative bank and a traditional bank investing in safe projects only absent cost of innovation (solid thick line) and three cost functions of innovation C_1, C_2 and C_3 (dashed lines).



First, consider any bank with a high cost of innovation $c(j) > \Delta_H$. In this case the cost to innovate is always greater than the increased profitability Δ from being an innovative bank, and this is true regardless of how many other firms innovate. In particular $c(j) > \Delta_H \geq \Delta|_{\mu=\hat{\mu}}$ for any $\hat{\mu}$ and so the benefits Δ from innovation will never cover the costs. It follows that any bank with a high cost of innovation (any bank j with $j > \mu$) will choose not to innovate.

Secondly, consider any bank j with a low cost of innovation $c(j) < \Delta_L$. In this case, the cost to innovate is always less than the increased profitability Δ from being an innovative bank, as long as the overall level of innovation μ is no greater than μ'. In particular, $c(j) < \Delta_L \leq \Delta|_{\mu=\hat{\mu}}$ for any $\hat{\mu} < \mu'$. We have already shown that bank j will not innovate if $j > \mu'$, and so it follows that the maximum level of innovation is indeed μ'. It follows that in equilibrium banks with a low cost of innovation (any bank j with $j \leq \mu$) will choose to innovate. This is sufficient to show that there exists a cost function such that there is a unique equilibrium where the level of innovation equals μ'. That is, all three scenarios are potential equilibria.

For the purpose of Proposition 4.4, we have assumed a particularly simple cost function. The set of cost functions that satisfy Proposition 4.4 will be large. Figure 2 shows three examples of cost functions (C_1, C_2 and C_3) that deliver different scenarios as the equilibrium.

5 Implications and extensions

In this section, we close the discussion with comments on the model and some simple extensions.

5.1 Alternative costs of information

5.1.1 Symmetric costs of information

Section 4 shows that any level of innovation below μ_m can be achieved as a unique equilibrium of the extended model when the cost of innovation is different for different banks (i.e. banks are ex-ante asymmetric). In this section, we consider the case where banks are ex-ante symmetric (i.e. all banks have the same cost of innovation). In this situation we have:

Proposition 5.1 *Suppose all banks have the same cost of innovation and $c(j) = c$ for all $j \in [0, 1]$. Then:*

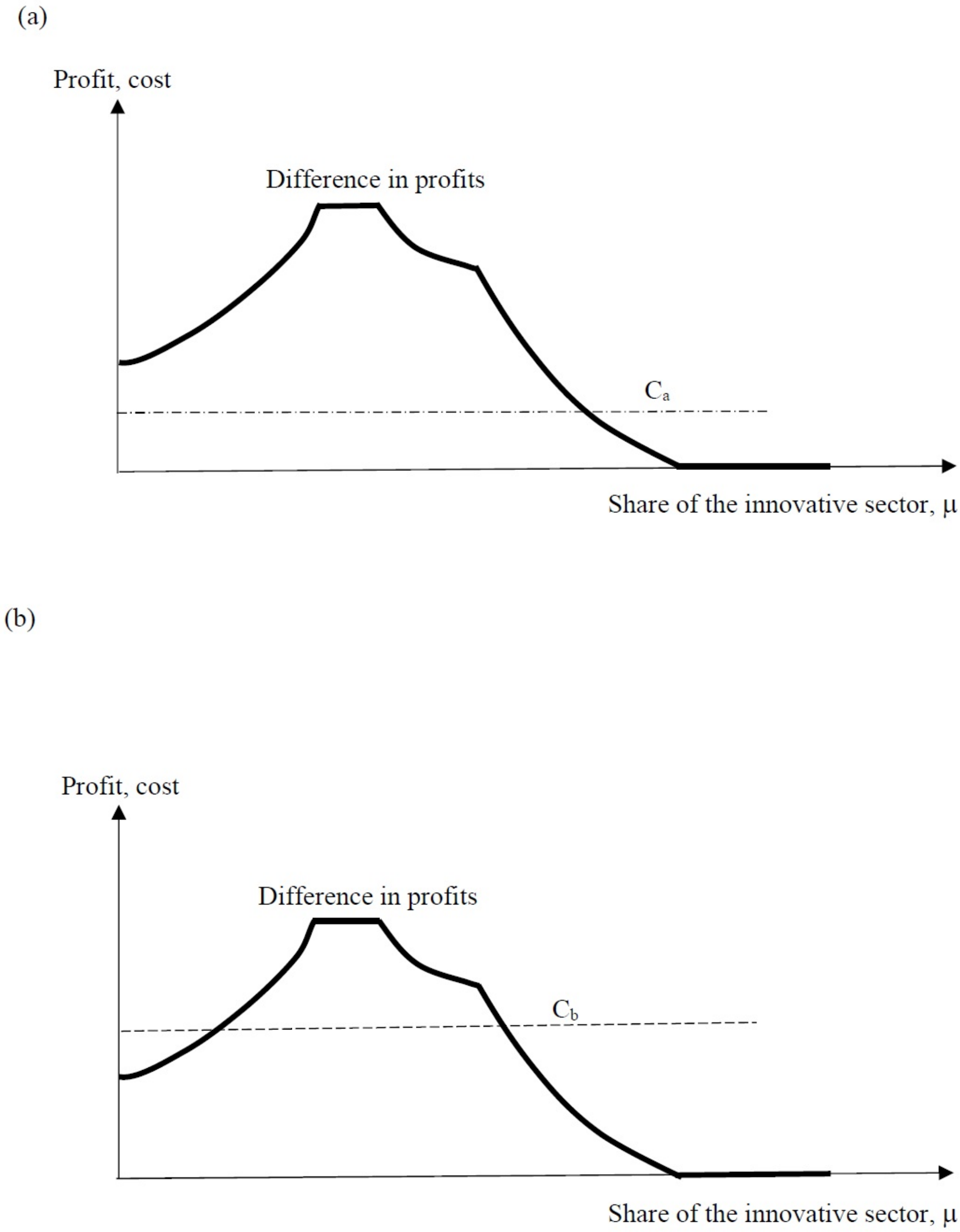
- *If $c \leq \Delta|_{\mu=0}$, then there is a unique equilibrium where the level of innovation μ is between μ^* and μ_m),*

- *If $c \geq \Delta|_{\mu=\mu^*}$, then there is a unique equilibrium where no bank innovates ($\mu = 0$),*

- *If $c \in \left(\Delta|_{\mu=0}, \Delta|_{\mu=\mu^*}\right)$, then there are two stable equilibria (i) where μ is between μ^* and μ_m, (ii) where no bank innovates ($\mu = 0$).*

The final part of this proposition considers the case when the cost of innovation is at an intermediate level and shows the possibility of multiple equilibria in the extended model. Multiple equilibria may arise because, in the increasing risk premia scenario, a larger number of innovative banks leads to an increase in the repayment rate risky projects pay, which boosts the profitability of existing innovative banks. This means that it may be the case that investing in the innovative technology is only profitable if other banks also decide to invest in the innovative technology (see Figure 3(a)). This creates a co-ordination problem, with two possible outcomes one where no firms innovate and one where the level of innovation is strictly positive. When the cost of innovation is sufficiently high, the extra profit earned by the innovative banks is never enough to cover the costs of innovation and hence no banks

Figure 3: Difference in profits between an innovative bank and a traditional bank investing in safe projects only absent cost of innovation (solid thick line) and a low constant cost function C_a (Panel (a)) and large constant cost function C_b (Panel (b)).



innovate. However, if the cost of innovation is sufficiently low, then innovation is profitable even if no other banks decide to innovate and in this case the unique equilibrium lies in the intermediate phase (between μ^* and μ_m) (see Figure 3(b)). We now turn to the case where the main cost of innovation lies not in investing in new technology but in purchasing data from a third party.

5.1.2 Cost to purchase data

Another interesting case arises if the cost lies not in the innovative technology itself but in the cost of the obtaining the data that is necessary to enable the innovative banks to separate out type A projects from type B projects. The market for data is interesting because much of the big data lies in the hands of very large players with monopoly power, notably, but not exclusively, the so called GAFAM (Google, Amazon, Facebook, Apple and Microsoft). Here we discuss a simple case where the owner of the relevant big data is a pure monopoly and charges an access fee to the data (here we are thinking of an ongoing fee related to usage, not a one-off access fee). Assuming that this is the only cost that innovative banks face, then the owner of the big data would set the access fee per innovative bank at a level that maximizes the aggregate profit (before subtracting access fees) of the innovative banks.

This set-up can be modeled in a similar way to the case above where all banks face the same cost c to innovate, but in this case the cost c (i.e. the access fee) is set endogenously by the data provider rather than being fixed exogenously. In particular, the data provider sets a fee c that has the effect of limiting the size of the innovative market. For example, define $\mu(c)$ as the unique equilibrium level of innovation when the cost to innovate equals $c \leq \Delta|_{\mu=0}$ and assume that the data provider sets the access fee $c \leq \Delta|_{\mu=0}$ to avoid the co-ordination problem discussed above and to maximize $\Pi(c) = c\mu(c)$.

Note, that when $c \leq \Delta|_{\mu=0}$, then $\mu(c)$ increases as c decreases, and hence the maximum charge that can be demanded from each innovative bank will be falling as the proportion of innovative banks increases. This results in the usual volume-price trade-off for the supplier of big data. We do not discuss this extension here, but an alternative approach could be for the data provider to offer a fee schedule that is conditional on μ. In this case, the equilibrium could be unique but with a fee potentially far higher with an associated smaller innovative sector. The general point from this section is that an external monopolistic supplier of data would set fees to restrict the scale of the innovative banking sector.

5.2 Imperfect identification of risky projects

So far we have assumed that innovative banks can perfectly distinguish between type A projects and type B projects. In this section we give an intuitive discussion of what happens if we relax this assumption. The section suggests that the core insights of the model with perfect identification of types of risky projects hold but with some interesting modifications.

The main change is that innovative banks are still likely, but now are not guaranteed, to prefer to lend to projects that they have 'identified' as type A rather than to the group of projects that are 'identified' as type B. de Roure *et al.* (2019) describe bottom fishing as the process of lending to risky, as opposed to safe, projects. In our model, innovative banks always lend to risky projects, so can be thought of as bottom fishing but, as shown in Section 4, they always lend to the less risky of the risky projects (i.e. lend to type A projects). However, if the innovative banks' identification procedure is poor and there is a large proportion of type B projects relative to type A projects, then it is possible that the innovative banks choose to lend to projects identified as type B, i.e. they choose to lend to the highest repayment risk borrowers.

To relax the assumption of perfect identification let us assume that when innovative banks assess whether risky projects are type A or type B their ML algorithms may incorrectly interpret information and mislabel the projects. We assume that the probability of error when classifying projects is ϵ. That is, with probability $\epsilon < 1/2$ all innovative banks erroneously label a type B project as A and vice versa.

Since $\epsilon < 1/2$, it follows that type B projects are more likely to be labelled B than A and vice versa. If ϵ is close to 0, then both type A and type B projects are almost certain to be labelled correctly and we are close to the special case of perfect identification analysed in Section 4. In this case (when ϵ is close to 0) it is straightforward to show that innovative banks continue to invest in projects labelled type A and we do not discuss this case further.

At the other extreme if ϵ is close to 1/2, then both type A and type B projects are labelled A and B with probability close to 1/2 and the ML technology's ability to identify risk is very weak. Recall that the overall proportion of risky projects that are actually type B projects equals $\zeta|_{\mu=0}$, and - given the ML technology's ability to identify risk is very weak - the proportion of risky projects that are actually type B projects will be very similar both (i) in the pool of projects labelled type A and (ii) in the pool of projects labelled type B. It follows that whether innovative banks target projects labelled type A or projects labelled type B they attract a very similar mix of type A and type B projects. In particular, if innovative banks choose to target projects labelled type A, then the proportion of risky projects served by innovative banks that are actually type B projects will be just less than $\zeta|_{\mu=0}$. Similarly, if innovative banks choose to target projects labelled type B, then the proportion of risky projects served by innovative banks that are actually type B projects will be just greater than $\zeta|_{\mu=0}$. This discussion motivates the following result:

Proposition 5.2 *Assume that $r_{A,F} > (1-q)s > r_{B,F}$. Then:*

- *If $\zeta|_{\mu=0} < \gamma_B$, then innovative banks invest in projects labelled type A whenever ϵ is sufficiently close to* $1/2$

- *If $\zeta|_{\mu=0} > \gamma_B$, then innovative banks invest in projects labelled type B whenever ϵ is sufficiently close to* $1/2$

The argument is similar to that of Section 4. To see this, first consider the case $\zeta|_{\mu=0} < \gamma_B$. Recall that when $\zeta|_{\mu=0} < \gamma_B$ and $\mu = 0$, the risk premium term $\mathcal{L}(r_R, \zeta)$ is increasing in ζ (as discussed in the increasing risk premia scenario). Since innovative banks always attract a similar mix of type A and type B projects compared to traditional banks, innovative banks choose to invest in projects labelled type A in order to decrease their exposure to type B projects and minimise the risk premium they face. However when $\zeta_0 > \gamma_B$ the risk premium term $\mathcal{L}(r_R, \zeta)$ is decreasing in ζ (as discussed in the intermediate scenario). In this case innovative banks choose to invest in projects labelled type B in order to increase their exposure to type B projects and to minimise the risk premium they face.

Finally, we close with a discussion of the impact of imperfect identification on risk premia. Under the case of perfect identification there is an increasing risk premia scenario as long as $\zeta|_{\mu=0} < \gamma_B$, but if $\zeta|_{\mu=0} > \gamma_B$ there is no increasing risk premia scenario. In contrast, if the ML technology's ability to identify differences in risk is poor, there is an increasing risk scenario whether $\zeta|_{\mu=0}$ is greater or less than γ_B. The reason is that for small positive μ traditional banks attract a very similar mix of type A and type B projects as they do when $\mu = 0$ (because the presence of innovative banks does not have a large impact on the mix of risky projects served by traditional banks). Whether innovative banks target projects labelled A (as happens when $\zeta|_{\mu=0} < \gamma_B$) or projects labelled B (as happens when $\zeta|_{\mu=0} > \gamma_B$), an increase in innovation will drive the mix of projects served by traditional banks (namely ζ) closer to γ_B. This increases the risk premium, which increases the repayment rates charged by traditional banks and innovative banks. This shows that introducing imperfect identification makes an increasing risk premia scenario more prevalent.

5.3 Competition, stability and concentrated business models

In this subsection we discuss the relationship between competition, concentrated business models and stability, an issue that has attracted considerable attention in the academic and policy literature, before we turn to policy implications in the following subsection. The traditional view is that greater competition for deposits is likely to increase instability since the higher deposit rates, induced by competition, encourage risk taking by banks. It is less clear-cut what happens if the greater competition arises on the asset side, which is the case considered in this paper. As noted in Section 2 greater competition for borrowers tends to reduce repayment rates, but the net effect on bank stability is ambiguous (see Boyd & De Nicolo (2005), Martinez-Miera & Repullo (2010)). Both Boyd & De Nicolo (2005) and Martinez-Miera & Repullo (2010) consider competition for borrowers from symmetrically positioned banks. In this paper, we analyse the impact of asymmetric forms of competition. The model provides insights into which factors matter and when, given that the impact of competition from innovative banks on the stability of the banking system depends on specific circumstances.

Instability in our model can arise from two core sources: (i) a change in the allocation of types of risky projects across banks and (ii) a change in the relative repayment rates. The emergence of innovative banks impacts the allocation of project types across innovative and traditional banks, and in particular innovative banks acquire disproportionately more of the lower repayment risk projects relative to their share in the economy.

Considering a change in relative repayment rates, other things being equal, a lower repayment rate for safe projects in conjunction with a higher repayment rate for risky projects will lead to more safe projects and less risky ones. It is clear from Figure 1 that in the increasing risk premia scenario, the change in relative repayment rates generates a 'flight to safety'. Thus, the change in relative repayment rates associated with innovation by banks tends to reduce aggregate risk taking by banks in the increasing risk premia scenario. It follows that if the parameters' values are such that at least one of the risky project types fails to cover repayments to depositors when the project fails, a change in relative repayment rates induced by the presence of innovative banks could be associated with greater financial stability in the banking sector. In general, the change in relative repayment rates tends to increase financial stability in the increasing risk premia scenario but tends to reduce it in the intermediate scenario.

The change in the allocation of risky projects across bank types, however, can offset the impact of the changes in relative repayment rates. An interesting example arises when repayments received from type A projects when they fail are greater than the deposit rate but type B projects cannot meet the repayments to depositors when they fail. In this case, if there are no innovative banks and the proportion of type A projects in the mix of risky projects is sufficiently high, then traditional banks always have more than enough money to repay depositors in all states of the world. Hence, with no innovative banks, the banking system is sound. We use the terminology general traditional banks to describe traditional banks with both risky and safe borrowers. In the face of small increases in the share of innovative banks, general traditional banks will still find that their portfolios of risky projects will, in aggregate, be able to repay depositors. However, at some point within the increasing risk premia scenario, there will be enough innovative banks that the return from the mix of the risky projects in general traditional banks will no longer cover the repayments to depositors if both types of the risky projects fail. Note that, general traditional banks will still be sound because there will be enough profit from the safe projects to make up the shortfall. However, general traditional banks have an alternative strategy. They could become banks with a concentrated business model, either concentrating on risky projects or concentrating on safe projects. There exists a critical proportion of innovative banks such that once the mix of risky projects in the general tradition banks is no longer able to meet the required return to depositors. At this point general traditional banks will be unable to compete against traditional banks with concentrated business models. The reason for this is deposit insurance. Any proportion of innovative banks beyond this critical proportion is associated with concentrated business models for all the banks. Hence, there is a critical proportion of innovative banks such that traditional banks lending solely to risky projects (we refer to such banks as risky traditional banks) and safe traditional banks (i.e. traditional banks lending solely to safe projects) dominate the traditional banking market from this point onward, and thus some risky traditional banks will fail in some states of the world. This generates instability in the banking system, even though general traditional banks would be solvent should they exist. Note, that this problem will happen whenever the higher repayment risk projects cannot cover repayments to depositors when they fail.

5.4 Policy options and tensions between micro and macro approaches to prudential regulation

In this section, we discuss policy implications. As background to the discussion, we begin by addressing a special case of the model that was not discussed in Section 4 and then address the welfare effects of innovative banking in the model.

The analysis in Section 4 is restricted to cases where the model parameters are such that traditional banks are willing to fund risky projects even if these are all type B projects. If this is not the case, then it is still possible that traditional banks will fund risky projects in the absence of innovative banks or, if projects of type B are particularly poor, then traditional banks will not fund any risky projects at all. In the latter case, innovative banks will be able to charge the highest repayment rates consistent with the supply of type A projects being sufficient to meet the demand for type A projects by innovative banks. The cost of innovation will determine how many banks adopt innovative technology. The impact of ML and the welfare implications are particularly simple in this case. Some type A projects, which previously were unable to obtain funding because traditional banks could not separate them from type B projects, can now be funded. This is clearly beneficial even though repayment rates for safe projects will be higher. This is a clear-cut case where ML can bring significant benefits but obviously this is an extreme case.

If the mix of type A and type B projects in the market is such that traditional banks will fund risky projects if there are no or few innovative banks, then the outcome will look like that in Section 4 providing that the proportion of innovative banks is not too large. However, if the proportion of innovative banks is large, traditional banks will not be willing to fund risky projects and type B projects will be driven from the market. The outcome will then be the same as the position in the above paragraph. The welfare effects when the mix of type A and type B projects in the market is such that traditional banks will fund risky projects if there are no or few innovative banks is very similar to the welfare effects of the model in Section 4, which we now consider.

A first point is that the impact on welfare that will arise from any change in the proportion of innovative banks will depend on the relative weights that are attached to the separate components of welfare. However, it is possible to provide some general comments. To begin with, it is very likely that in the increasing risk premia scenario a greater proportion of innovative banks leads to a drop in welfare. The clearest case arises if traditional banks are solvent even when both types of risky projects fail. At this point, there is no call on deposit insurance arising from a small increase in the share of innovative banks. However, the repayment rate charged to risky projects is not based on the actual mix of low and high repayment risk projects in the economy, but instead is distorted above this level by the presence of innovative banks. Hence, the trade-off between risky and safe projects is disproportionately skewed in favour of safe projects, creating a welfare loss (because at this point neither risky nor safe projects cause a financial problem for any banks). This type of argument causes a welfare 'problem' all through the increasing risk premia scenario. However, once one gets to the segregated scenario, the repayment rates of the lower repayment risk projects fall and the problems begin to unwind, suggesting that at this point welfare is likely to increase with a larger share of innovative banks.

If the proportion of innovative banks is close to μ_m, both type A and type B projects are paying repayment rates that are close to those that would arise if there were full information. However, whether welfare is higher at μ close to μ_m, or without any innovative sector (i.e. $\mu = 0$) depends on two counter effects. On the one hand, without any innovative sector both types of risky projects pay the same rate, so there are too many high repayment risk projects relative to low repayment risk projects. It is possible at this point ($\mu = 0$) that traditional banks are always solvent, so in this case, there would be no moral hazard problem. It follows that, save for not being able to separate types of risky projects, there is the right mix of risky and safe projects. On the other hand, when μ is close to μ_m, we have almost the right allocation between the types of risky projects, but we also have a moral hazard problem induced by the deposit insurance (i.e. banks finance too many risky projects). Welfare is higher with a large innovative sector (with μ close to μ_m) rather than with no innovative sector (with $\mu = 0$) when the welfare loss caused by the moral hazard problem generated by deposit insurance is smaller than the welfare loss caused by the mis-allocation of funds due to banks being unable to distinguish type A projects from type B projects. If welfare is greater at μ close to μ_m, then the government may wish to promote innovation in banks, whereas if the welfare is greater without innovative banks, the government may wish to discourage the innovative sector (or to be more precise, discourage those fintech banks whose only benefit is the ability to separate between project types).

We now discuss capital requirements because they are the sole policy instrument in the basic model. We should point out, however, that there are other possible instruments that would probably be better at resolving the problem without recourse to capital requirements, although many of these would be outside the scope of a prudential regulator. We return to this point following the discussion of capital requirements.

We assume that the opportunity cost of capital is greater than the cost of deposits.[9] As a result, capital is relatively expensive and banks will choose no more capital than the minimum ratio, q, set by the prudential regulator. Increasing q, if all banks face the same q, typically improves stability. Holding other parameter values constant, an increase in q would shift the point where traditional banks must adopt concentrated business models to survive and, hence, increases the range of μ for the increasing risk premia scenario in which all the banks are stable. Generally, increasing q makes risky projects less attractive to banks for the standard reason (i.e. since more of the bank's money is at risk), which leads to lower repayment rates for safe projects. Thus, there are more safe projects and greater stability.

Given the diversity of risk profiles across bank types, it is more likely that differential rates for different types of banks will provide greater policy flexibility. The policy choices and tensions are particularly interesting if $r_{A,F} > (1-q)s > r_{B,F}$, and the remainder of this section assumes this special case. If this assumption holds, neither traditional safe banks nor innovative banks fail in any state of the world, but traditional risky banks may or may not fail when both types of risky projects fail. Whether they fail or not depends on the mix of risky projects they hold. Policy outcomes will depend on which scenario (i.e. increasing risk premia, intermediate or separated) the market is in.

If the costs of innovation are such that the equilibrium is in the segregated scenario, then neither safe traditional banks nor innovative banks fail but risky traditional banks fail whenever type B projects fail. The natural policy response of raising capital requirements for risky traditional banks (but not for safe traditional banks or innovative banks) will lead to a rise of the repayment rates offered by risky traditional banks. The repayment rates offered by innovative and safe traditional banks will decline (since the market clearing condition implies some repayment rates must fall if others rise). Consequently, there will be a shift from risky traditional banks to innovative and safe traditional banks leading to an increase in stability.

If the costs of innovation are such that the equilibrium is in the increasing risk premia scenario, then the policy alternatives are less straightforward. Consider first the effect of

[9]One could think of this arising because bank owners have private opportunities available to them that earn a rate greater than the opportunities that are available to depositors.

increasing capital requirements for innovative banks and not for traditional banks. At the existing repayment rates the marginal innovative bank will now lose money and so the equilibrium requires less innovative banks. However, less innovative banks will improve the mix of risky projects in the traditional banks' portfolios and hence the repayment rate for risky projects will fall. Consequently, the equilibrium repayment rate for safe projects must increase. Therefore, an increase in capital requirement for innovative banks only, would lead to a fall in the repayment rate of risky projects with the cost of the capital requirements being passed on to safe borrowers despite safe traditional banks facing no increase in capital requirement. The impact on stability depends on the mix of projects in risky traditional banks. If, after the increase in capital requirement, risky traditional banks fail when type B projects fail, then there is a decline in stability. That is, the change in capital requirement has led to a smaller proportion of those banks who do not fail in any state of the world and a larger proportion of those banks that fail in the state of the world when both risky projects fail. However, if the increase in the proportion of type A projects in risky traditional banks is large enough to ensure that these banks do not fail when both types of projects fail, then this results in an increase in stability. This latter case leads to a potential tension between micro and macro approaches to prudential regulation, which we now address.

The tension is easiest to elucidate in the following special case. Suppose there exists $\hat{\mu} > 0$ such that risky traditional banks can just repay depositors when $\mu = \hat{\mu}$ and that the costs of innovation are such that the equilibrium μ is slightly greater than $\hat{\mu}$. Hence, risky traditional banks fail when both types of risky projects fail. As discussed above, a small increase in capital requirement for innovative banks alone will return the system to full stability. Note, an increase in capital requirement for risky traditional banks alone may also return the system to full stability. However, an increase in capital requirement for risky traditional banks alone raises risky repayment rates. The increase in risky repayment rates, in turn, increases the share of innovative banks, which reduces the proportion of type A projects within risky traditional banks' portfolios. This further increases risky repayment rates since the increase has to offset the reduction in type A projects and the increase in the higher capital requirement. If the slope of the cost of innovation as a function of μ is almost vertical, then the feedback loop problem from the increase in the share of innovative banks would be small and so the increase in capital requirement to return traditional banks to stability might not need to be too large. However, it is possible that the slope of the cost function is such that an increase in repayment rate leads to a large increase in μ and the capital requirement needed to return risky to full stability is very large or may not even exist.

A micro prudential approach will focus on capital requirements for risky traditional banks and may require very high capital requirements, which may themselves fail to lead to a return to full stability. In contrast, a macro prudential approach would consider the best policy to reduce risk in the system as a whole and as such would probably opt for raising capital requirements for innovate banks alone. While this may appear to be an odd approach, it is in one sense simply a reflection of the underlying adverse selection problem and the standard polluter pays principle applied to it (i.e. taxing the agent creating the externality that is causing the underlying welfare issue).

We have considered capital requirements in some detail because they are the sole policy instrument in the basic model but, as indicated earlier, broader policy tools could also apply. If costs are such that the equilibrium is in the segregation scenario, then policies that amount to a subsidy for innovation are beneficial. This could arise through favourable treatment of innovation or through disadvantaging risky traditional banks. In each case there is no natural tension between macro and micro prudential approaches. In contrast, if costs of innovation are high and the equilibrium is in the increasing risk premia scenario, then a limited intervention that restricts entry may improve stability and increase welfare, whereas actions to disadvantage risky traditional banks may be difficult and even ultimately unsuccessful. Hence, the tension between macro and micro prudential policy is also likely to be present when other policy instruments are available.

6 Conclusions

There appears to be a consensus that ML and Big Data will impact the banking sector but there are many differing views as to what this impact will be. One message of our paper is that there is no unique answer to the question 'what will the impact be?'. Specifically, the paper considers the impact of ML in credit markets where those adopting ML (innovative banks) compete with traditional banks (those banks that have not adopted ML) and shows that the impacts of ML in credit markets differ dramatically according to the characteristics of the market. However, the core message of the paper is that, despite the varied outcomes, it is possible to elucidate clear and simple relationships between characteristics and impacts.

For example, among other results, we show that if (i) the cost of adopting ML technology is high, (ii) there is a small proportion of higher repayment risk projects among the set of risky projects and (iii) the ML technology is good at identifying types of risk projects, then innovative banks will target the low repayment risk projects but there will be weak competition among innovative banks for low repayment risk projects. As a result, the presence of innovative banks raises repayment rates for all risky projects (high and low risk), i.e. the low repayment risk borrowers lose out from the presence of innovative banks, despite being targeted by the innovative banks. However, the repayment rates for safe borrowers will be lower (i.e. there is a flight to safety).

In contrast, if (i) the cost of adopting ML technology is low, (ii) there is a small proportion of low repayment risk projects among the set of risky projects and (iii) the ML technology is good at identifying types of risky projects, then innovative banks will again target the low repayment risk borrowers but there will be strong competition among innovative banks for low repayment risk borrowers. This will bring lower repayment rates for low repayment risk projects (bringing more of them into the market) and traditional banks will have concentrated business models, some specialising in lending to safe borrowers and others lending only to the highest repayment risk borrowers in the market. An extreme version of this applies if the high repayment risk borrowers are not viable if not pooled with low repayment risk borrowers. In this case, the innovative banks will drive the high repayment risk borrowers from the market.

In the examples above, innovative banks adopt a cream skimming strategy. Indeed, we show that this is the case whenever the ML technology is good at identifying types of risky projects. However, cream skimming may not be the best policy if ML technology is not accurate in identifying types of risky projects. Specifically, if the costs of adopting ML technology is high and there is a small proportion of high repayment risk projects among the set of risky projects, and the ML technology is very poor at identifying types of risky projects, then innovative banks will target the high repayment risk projects (bottom fishing) and the repayment rates for all risky borrowers (high and low risk) will be higher than in the absence of innovative banks.

Regardless of whether the ML technology is good or poor at identifying differences in risk, we also show that if the cost of adopting ML is relatively similar across all banks, then multiple equilibria is the likely outcome, with network effects potentially preventing the development of innovation.

Given that different characteristics of the market produce very different outcomes, it is not surprising that the effect of alternative policy approaches also differs enormously according to the exact nature of the market. Of the examples we have given above, the second one is the situation where ML brings significant benefits and is clearly the end-game that is hoped for. However, as pointed out, the presence of innovative banks pushes traditional banks towards concentrated business models. In this situation a conventional policy response to aid innovative banks whilst 'restricting' risky traditional banks (e.g. a subsidy for innovation and an increase in capital requirements for traditional banks specialising in risky borrowers) will increase stability and welfare. This is in stark contrast to the first example, where, due to the high cost, innovation has a more niche share of the market. In this case conventional policies have unexpected outcomes. For example, an increase in capital requirements for innovative banks only (or a similar policy choice) would reduce repayment rates for all risky borrowers rather than increase them, and the cost of the capital requirements will be paid by the safe borrowers through higher repayment rates. Furthermore, as discussed, this leads to a tension between policies aimed at reducing risk in the whole system (macroprudential policy) and those aimed directly at individual risky banks (microprudential policy).

References

Albanesi, S., & Vamossy, D.F. 2019. Predicting consumer default: A deep learning approach. *NBER Working Paper 26165*.

Auer, R., Cornelli, G., & Frost, J. 2020. Rise of the central bank digital currencies: drivers, approaches and technologies. *BIS WP 880*.

Ayygai, M., Demirgüç-Kunt, A., & Maksimovic, V. 2010. Formal versus informal finance: Evidence from China. *Review of Financial Studies*, **23(8)**, 3048–3097.

Balyuk, T., Berger, A.N., & Hackney, J. 2020. What is fueling FinTech lending? The role of banking market structure. *SSRN Working Paper*.

(BBB), British Business Bank. 2019. 2019 Business Finance Survey: SMEs.

Beck, T., Da-Rocha-Lopes, S., & Silva, A.F. 2020. Sharing the pain? Credit supply and the real effect of banks bail-ins. *Review of Financial Studies*, **in press**.

Behr, P., Norden, L., & Noth, F. 2012. Financial constraints of private firms and bank lending behavior. *Journal of Banking and Finance*, **37(9)**, 3472–3485.

Bellucci, A., A., Borisov, & Zazzaro, A. 2010. Does gender matter in bank–firm relationships? Evidence from small business lending. *Journal of Banking and Finance*, **34(12)**, 2968–2984.

Berg, T. 2018. Got rejected? Real effects of not getting a loan. *Review of Financial Studies*, **31(12)**, 4912–4957.

Berg, T., Burg, V., Gombović, A., & Manju, P. 2020. On the rise of Fintechs: Credit scoring using digital footprints. *Review of Financial Studies*, **33**, 2845–2897.

Bolton, P., Freixas, X., Gambacorta, L., & Mistrulli, P.E. 2016. Relational and transaction lending in a crisis. *Review of Financial Studies*, **29(10)**, 2643–2676.

Boot, A.W.A., & Thakor, A. 2002. Can relationship banking survive competition? **55(2)**, 679–713.

Boyd, J.H., & De Nicolo, G. 2005. The Theory of Bank Risk Taking and Competition Revisited. *J. Finance*, **60**, 1329–1343.

Canales, R., & Nanda, R. 2012. A darker side to decentralized banks: Market power and credit rationing in SME lending. *Journal of Financial Economics*, **105(2)**, 353–366.

Chen, M.A., Wu, Q., & Yang, B. 2019. How valuable is FinTech innovation? *Review of Financial Studies*, **32(5)**, 2062–2106.

Claessens, S., Frost, J., Turner, G., & Zhu, F. 2018. Fintech credit markets around the world: size, drivers and policy issues. *BIS Quarterly Review*.

Crawford, G.S., Pavanini, N., & Schivardi, F. 2018. Asymmetric information and competition in lending markets. *The American Economic Review*, **108(7)**, 1659–1701.

Cumming, D. 2008. Contracts and exits in Venture Capital finance. *Review of Financial Studies*, **21(5)**, 1947–1982.

de Roure, C., L., Pelizzon, & A.V., Thakor. 2019. P2P lenders versus banks: Cream skimming or bottom fishing? *SAFE Working Paper 206*.

Di Maggio, M., & Yao, V. 2020. Fintech borrowers: Lax-screening or cream-skimming? *NBER Working Paper 28021*.

Dorfleitner, G., Priberny, C., Shuster, S., J., Stoiber, M., Weber, de Castro, I., & Kammler, J. 2016. Description-text related soft information in peer-to-peer lending – Evidence from two leading European platforms. *Journal of Banking and Finance*, **64**, 168–187.

Fay, M., & Williams, S. 1993. Gender bias and the availability of business loans. *Journal of Business Venturing*, **8(4)**, 363–376.

Fernández-Delgado, M., Cernadas, E., Barro, S., & D., Amorim. 2014. Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, **15**, 3133–3181.

Freedman, S., & Jin, G.Z. 2017. The information value of online social networks: Lessons from peer-to-peer lending. *International Journal of Industrial Organisation*, **69**, 185–222.

Frost, J. 2020. The economic forces driving fintech adoption across countries. *BIS WP 838*.

Frost, J., Gambacorta, L., Huang, Y., S., Shin H., & P., Zbinden. 2019. BigTech and the changing structure of financial intermediation. *Economic Policy*, **forthcoming**.

(FSB), Financial Stability Board. 2019. FinTech and market structure in financial services: Market developments and potential financial stability implications.

Fu, R., Huang, Y., & Singh, P. 2020. Crowd, lending, machine and bias. *SSRN Working Paper*.

Fuster, A., Plosser, M., Schnabl, P., & Vickery, J. 2019. The role of technology in mortgage lending. *Review of Financial Studies*, **32**, 1854–1899.

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., & Walther, A. 2020. Predictably unequal? The effects of machine learning on credit markets. *SSRN Working Paper.*

Galli, E., Mascia, D.V., & Rossi, S.P.S. 2020. Bank credit constraints for women-led SMEs: Self-restraint or lender bias? *European Financial Management*, **26(4)**, 1147–1188.

Gambacorta, L., Huang, Y., Qiu, H., & Wang, J. 2019. How do machine learning and non-traditional data affect credit scoring? New evidence from a Chinese fintech firm. *BIS Working Papers No 834.*

Goel, A., Song, F., & Thakor, A.V. 2021. Banking on hard and soft information. *Washington University at St. Louis Working Paper.*

Granja, J., Leuz, C., & Rajan, R.G. 2018. Going the extra mile: Distant lending and credit cycles. *NBER Working Paper 25196.*

Hand, D.J. 2006. Classifier technology and the illusion of progress. *Statistical Science*, **21(1)**, 1–14.

Hazelmann, R., & Wachtel, P. 2010. Institutions and bank behavior: Legal environment, legal perception, and the composition of bank lending. *Journal of Money, Credit and Banking*, **42(5)**, 965–984.

Iyer, R., Khwaja, A. I., Luttmer, E. F., & Shue, K. 2016. Screening peers softly: Inferring the quality of small borrowers. *Management Science*, **62**, 1554–1577.

Jagtiani, J., & Lemieux, C. 2019. The roles of alternative data and machine learning in Fintech lending: Evidence from the LendingClub consumer platform. *Financial Management*, **48(4)**, 1009–1029.

Karolyi, S.A. 2018. Personal lending relationships. *Journal of Finance*, **73(1)**, 5–49.

Kersten, R., Harms, J., Liket, K., & Maas, K. 2017. Small firms, large Impact? A systematic review of the SME finance literature. *Word Development*, **97**, 330–348.

Lessmann, S., Baesens, B., Seow, H-V., & Thomas, L.C. 2015. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, **247**, 124–136.

Martinez-Miera, D., & Repullo, R. 2010. Does Competition Reduce the Risk of Bank Failure? *The Review of Financial Studies*, **23**, 3638–3664.

Milone, M. 2019. Smart lending. *https://www.mariomilone.org/publication/smart-lending/.*

Ongena, S., & Popov, A. 2016. Gender bias and credit access. *Journal of Money, Credit and Banking*, **48(8)**, 1691–1724.

Padhi, M. 2017. The effects of peer-to-peer (P2P) lending on competition, discrimination, and financial stability. *University of Maryland Working Paper.*

Palia, D. 2016. Differential access to capital from financial institutions by minority entrepreneurs. *Journal of Empirical Legal Studies*, **13(4)**, 756–785.

Parlour, C.A., Rajan, U., & Zhu, H. 2020. When FinTech competes for payment flows. *SSRN Working Paper 3544981.*

Philippon, T. 2016. The Fintech Opportunity. *NBER Working Ppaer 22476.*

Puri, M., Rocholl, J., & Steffen, S. 2017. What Do a Million Observations Have to Say about Loan Defaults? Opening the Black Box of Relationships. *Journal of Financial Intermediation*, **31**, 1–15.

Riley, J.G. 1979. Informational equilibrium. *Econometrica*, **47(2)**, 331–359.

Rodano, G., Serrano-Velarde, N., & Tarantino, E. 2018. Lending standards over the credit cycle. *Review of Financial Studies*, **31(8)**, 2943–2982.

Tang, T.T. 2009. Information asymmetry and firms' credit market access: Evidence from Moody's credit rating format refinement. *Journal of Financial Economics*, **93**, 325–351.

Vallée, B., & Zeng, Y. 2019. Marketplace lending: A new banking paradigm? *Review of Financial Studies*, **32(5)**, 1939–1982.