The impact of machine learning and big data on credit markets

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

[1]CEPR, [2]Bank of England, [3]University of Bristol, [4]University of Bath

September 2020

Abstract

There is evidence that machine learning (ML), often using Big Data, can improve the screening of risky borrowers, and hence, potentially improve the flow of lending to smaller business entities, e.g. small and medium sized enterprises. This paper explores a scenario where traditional banks compete with fintech (innovative) banks that have these new technologies. The paper shows that there are significant non-linearities associated with the entry of innovative banks, which makes for complex impacts on credit markets, stability, and policy. The core idea is that at the emerging phase, the dominant form of competition is that between the traditional banks. At this point, by distorting the mix of risky projects within traditional banks, innovative entries create an externality for all risky borrowers, raising repayment rates for risky projects, choking off demand for loans and leading to a 'flight to safety'. Once innovative banks become a 'significant' share of the market, competition by innovative banks becomes a central driver on prices and the impact of innovative banks on the traditional banks mix of projects is less important. The desire to suck more lower risk projects into the market begins to drive down repayment rates for lower risk borrowers and the negative externality starts to unravel. The paper models these effects and discusses implications for competition, stability, concentrated business models and the market for Big Data.

JEL-Classification: G21, G28, G32, G28, 031, 033

Keywords: banking, credit markets, Big Data, machine learning, fintech, adverse selection, capital requirements, prudential regulation

1. Introduction

The use of information technology in financial markets and within financial institutions, fintech, is one of the most interesting, and potentially game changing developments in recent years. Its impact is likely to be felt widely throughout the sector, including regulatory agencies (regtech). One development, still at a relatively nascent stage but with potential for rapid growth and major impact, is the use of machine learning (ML), often using Big Data, to better identify risks and differences in risks between those seeking credit. It is particularly hoped that it will help improve credit flows to smaller business entities, e.g. small and medium enterprises (SMEs), that often struggle to obtain good 'deals' from traditional banks but play an important role in innovation and an economy's growth. The ability of fintech innovations, both to increase the scale of information gathered and its interpretation, has been debated by academics (e.g. Hand, 2004; Fernández-Delgado et al. 2014; Lessmann et al., 2015) and regulators (BoE 2019, 2020; Auer et al. 2020; Frost, 2020). Yet little is known about how the implementation of fintech solutions will affect the structure of the financial markets and the quality of the services offered. This paper contributes to the debate by discussing the impact a more accurate assessment of borrowers' risk profile will have on the repayment rates they are offered by traditional and innovative banks.

According to the World Bank, SMEs account for about 90% of businesses and more than 50% of employment worldwide.[1] Moreover, the World Bank estimates that 600 million new jobs will need to be created by 2030 to meet needs of the growing workforce, and a significant proportion of these jobs will need to be created by SMEs. However, regardless of the importance of SMEs for economic growth and fighting poverty, SMEs face greater difficulties than big corporations in accessing funding to finance their activities and growth. These difficulties in accessing formal funding are not restricted to developing countries. In the UK, for example, 30% of SMEs report that they face difficulties in accessing the funding they need.[2] Only 10% of them used bank loans and commercial mortgages. A further 17% used credit cards as a source of financing. Bank overdrafts were the most common source of financing and accounted to 22% of all funding used by SMEs.[3] Even though only 8% succeeded in securing bank loans, this still amounts to a significant amount of money. According to UK Finance, the

[1] https://www.worldbank.org/en/topic/smefinance
[2] https://www.merchantsavvy.co.uk/uk-sme-data-stats-charts/
[3] https://www.bva-bdrc.com/products/sme-finance-monitor/

gross new loans obtained by SMEs amounted to £57 billion in 2017 and constituted nearly 68% of all gross new funding.[4]

SMEs face a similarly dire situation in the U.S. where, in 2015, 63% of companies with revenue of less than $100,000 and 58% of startups reported not being able to obtain required finances.[5] According to the more recent survey, in 2018, 32.7% of U.S. SMEs' loan applications were rejected while as many as 55.2% of them applied for loans (OECD 2020).

The adoption of much improved data predictive analytics, thanks to innovations related to the availability of Big Data and the use of ML methodologies, has the potential to dramatically improve the screening risky borrowers. The literature so far has primarily focused on the role of non-bank fintech lenders, as opposed to banks that rely on more traditional screening methodologies based on proprietary soft and hard information gathered over time through the course of their lending relationships. There is emerging evidence that the adoption of ML innovations can have positive effects in terms of identifying projects with lower default risk (Iyer et al., 2015; Dorfleitner et al., 2016; Berg et al., 2019; Frost et al., 2019; Fuster et al., 2019; Gambacorta et al., 2019).

It is too early to claim that a consensus has been reached on whether it is the case that non-bank fintech lenders can systematically outperform banks (Freedman and Jin, 2017; and Claessens et al., 2018). Nevertheless, the adoption of these innovative methodologies by fintech banks should, in principle, add to their ability to manage credit risk and compete with more traditional banks. This paper focusses on the competition dynamics between traditional banks and fintech banks that have these new technologies.

The paper addresses the impact of ML on credit markets at different stages of the development of the innovative market, from a small emerging sector to a mature one. The core feature of the paper is to show that there are significant non-linearities associated with the entry of innovative banks, i.e. as the innovative sector grows the impact on the credit market is not simply a bigger version of the impact with less innovation. Indeed, small levels of entry can have the exact opposite effect to larger levels of entry. This makes for complex impacts on credit markets, stability, and policy.

[4] https://www.ukfinance.org.uk/system/files/UK-Finance-SME-Finance-in-UK-AW-web.pdf
[5] https://www.pymnts.com/news/b2b-payments/2016/smes-severely-dissatisfied-with-elending-experience-fed/

Our results are mainly driven by one factor and the consequences that follow from it. Namely, if there is only a small number of innovative players in a market and they are able to 'cream off' a small proportion of the borrowers with lower risk projects, then repayment rates will still be driven by the competition between the traditional banks that service all types of borrowers with risky and safe projects. Thus, the main effect of innovative banks' entry arises from the impact of the innovative banks on the quality of the borrowers with risky projects that remain with the traditional banks, and the effect that this has on the traditional banks competing for borrowers with safe and risky projects. The pool of borrowers with risky projects remaining with the traditional banks is slightly worse on average, so the traditional banks will offer borrowers with risky projects a worse deal, relative to borrowers with safe projects, than before the innovative banks entered. Hence, the entry of the innovative banks creates a negative externality that makes things worse for all borrowers with risky projects and creates a flight to safety. Once innovative banks gain a 'significant' share of the market, then competition by the innovative banks for the borrowers with lower risk projects starts to bite and impacts the repayment rates. The desire to suck more borrowers with lower risk projects into the market begins to drive down repayment rates for the borrowers with lower risk projects and raises rates for the borrowers with higher risk projects, undoing the problems generated when the innovative sector was small. The main point is that the size of the innovative sector does not have a linear effect. A small innovative sector can create negative impact for all risky borrowers, but a large innovative sector can create significant benefits in the right situation.

To focus attention on the implications of the effect described above, the model in the paper takes a simple approach to capture the difference between traditional and innovative banks. We assume that there are three types of borrowers: those with safe projects and two types of borrowers with risky projects, one better than the other. While traditional banks can observe the difference between borrowers with safe and with risky projects, they are unable to identify the difference between borrowers with lower and higher risk projects. As indicated above, innovative banks can more accurately identify the type of the risky borrower, and for simplicity we take the extreme case that they are able to perfectly identify the difference between borrowers with low and high-risk projects. The core insights of the paper do not depend on the assumption that innovative banks can perfectly identify borrowers with low risk projects. The central point is that they have the ability to identify borrowers with low risk projects more accurately than traditional banks.[6]

The presence of a small number of innovative banks in the market sucks off some of the borrowers with low risk projects, since the innovative companies have no interest in taking any of the borrowers with high risk projects. Competition between the traditional banks will ensure that, in most circumstances, the repayment rates offered to borrowers with high risk projects by traditional banks will increase, because the risk of the average borrower with a risky project using traditional banks has become worse. The model assumes that the lower the repayment rates offered to specific types of borrowers, then the more of that type will become available. The higher repayment rates offered to borrowers with risky projects will curtail their interest in borrowing, and the traditional banks will have to look to encourage more borrowers with safe projects to fill the gap. This means that the repayment rates asked from borrowers with safe projects will have to fall. Hence, there will be a 'flight to safety' as traditional banks build up the proportion of borrowers with safe projects in their banking books. In this case innovative banks, in aggregate, are small in the marketplace so will only be able to service a relatively small number of borrowers with low risk projects, hence to find enough borrowers the innovative banks only need to match what the borrowers with low risk projects can get elsewhere. So, the presence of a small number of innovative banks has the opposite effect to what one might anticipate. The deal offered to borrowers with risky projects gets worse and the deal offered to borrowers with safe projects gets better. Borrowers with risky projects are discouraged while borrows with safe projects are encouraged, so risky projects (both the higher and lower risk projects) get squeezed out and, generally, aggregate stability in the sector increases.

While the innovative sector is in the emerging stage, the more innovative banks there are, then the bigger the externality and the greater the scale of the problems identified above. However, eventually there will be so few lower risk borrowers served by traditional banks that, even when the lower risk projects are successful, traditional banks cannot repay depositors when the high-risk projects fail. At this point (two) intermediate phases can emerge. The repayment rates for the borrowers with risky projects no longer increase and then start to fall, at which point

[6] In particular, the four stages and the relationship between repayment rates in each stage identified in Section2 will be preserved if innovative banks can only observe an imperfect signal of a borrower's type.

the negative externality begins to unravel. As the innovative sector grows, at some point, it will become harder for all the innovative banks to find enough of the lower risk borrowers. Competition between the innovative banks for lower risk borrowers starts to intensify and innovative banks will begin to undercut each other. The repayment rates offered to the lower risk borrowers will fall, which will bring more borrowers with low risk projects into the market. To accommodate the increasing number of borrowers with low risk projects the repayment rates offered by traditional banks to their risky borrowers (these will by now made up entirely of the more risky projects) and their safe projects will go up. The more innovative banks there are, then the more the externality problems produced in the emerging stage is unwound. The flight to safety starts to unravel and the instability of the banking sector increases.

As the proportion of innovative banks in the market increases, then eventually the only way to bring more of the lower risk borrowers into the market is to offer a repayment rate that is economically unattractive. At this point, in the model, there is no benefit to having more innovative banks in the marketplace. Obviously, in a model that focuses on other features of innovation, for example, lower overall cost of delivery, there may be scope for further innovative entry but in our model further innovation brings no benefit to the marketplace. In the remainder of the paper we model the effects identified above and then discuss the implications and extensions.

The following section outlines the model and presents the main results. Section Three discusses extensions and the implications for competition, stability, concentrated business models and the market for big data.

2. Model

Let us 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. All projects of a given type are identical. Safe projects S never fail, i.e. the borrowers always repay the money borrowed from banks. However, both type A and type B projects may succeed or fail. More precisely, 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 at the same time equals p_{AB}. By construction $p_{AB} < p_A$ and $p_{AB} < p_B$. When type A projects fail, they repay r_A^L . Analogously, when type B projects fail, they repay r_B^L. When projects succeed banks are repaid in full.

A proportion μ of total funds held by banks are held by innovative banks, while the remaining funds are held by traditional banks. Traditional banks can identify safe projects but cannot distinguish between type A and type B risky projects. Innovative banks can distinguish between all project types. Each bank chooses a repayment rate to offer each type of project. In particular, banks offer repayment rates r_{θ}^{i}, where $i \in \{I, T\}$ 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. If traditional banks cannot distinguish between the type of the risky projects, a repayment rate is denoted as r_{R}^{T}.

We also assume that type A projects have a more attractive return profile than type B projects. We assume that the following conditions hold:

$$\mathrm{r}_A^L \geq \mathrm{r}_B^L \quad (1)$$

$$p_A \leq p_B \quad (2)$$

$$p_A/\mathrm{r}_A^L < p_B/\mathrm{r}_B^L \text{ , i.e. one of the above inequalities is strict.} \quad (3)$$

For simplicity of notation, we refer to type A projects as the low-risk projects and project B as the high-risk projects.

We also assume that banks have a total quantity of funds D to allocate to the projects. We also assume that for each bank, a proportion of funds, q, must be provided by capital investors while remaining funds are provided by retail depositors. Capital investors can either invest in banks or invest in an outside option with expected return c . Retail depositors are always paid a deposit rate s , which we assume is less than c . 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.

The aggregate demand for loans $\Phi_S(\cdot)$, $\Phi_A(\cdot)$ and $\Phi_B(\cdot)$ from projects of type S, A and B respectively, depends on the competitive repayment rates r_S, r_A, r_B and r_R (if banks do not observe differences between type A and type B projects). We assume that all the demand functions, i.e. Φ_S, Φ_A, Φ_B, are strictly decreasing. This captures the fact that 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 $\varsigma \in [0,\ 1]$.

2.1. Stages of development of innovative banks

In this sub-section we consider four stages of development of innovative banks. The first stage we call the emerging stage, when innovative banks have a relatively small share of the market and there is a single repayment rate for all risky projects. In this stage, the repayment rate for risky projects increases and for safe projects decreases as the size of the innovative sector increases. We then have two intermediate stages. The first is called the unsegregated intermediate stage. In this stage, both type *A* and type *B* projects are held by traditional banks. There is a single repayment rate for all risky projects and the repayment rate for risky projects is weakly decreasing as the innovative sector increases. The repayment rate for safe projects is weakly increasing. The second intermediate stage is called the segregated intermediate stage. In this stage traditional banks do not hold type *A* projects and there are separate repayment rates for each of the project types. In this stage the repayment rates for type *A* projects decline as the innovative sector grows, whilst the repayment rates for safe and type B projects increase. We refer to the final stage as the mature stage. In this stage, the repayment rates are constant.

The thresholds between the four stages are denoted:

- μ^* - the boundary between the emerging and the unsegregated intermediate stages,
- μ^{**}- the boundary between the unsegregated and the segregated intermediate stages, and
- μ^{***} - the boundary between the segregated intermediate and the mature stages.

However, it is useful, before we discuss the three stages, to address the situation when there are no innovative banks.

2.1.1. 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 both type A and type B projects are the same, and denoted r_R^T. These 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. That is,

$$\begin{aligned} r_S^T = (1 - p_A - p_B + p_{AB}) r_R^T + (p_A - p_{AB})\, max\{(1-q)s, (1-\varsigma(0))\mathrm{r}_A^L + \varsigma(0) r_R^T\} \\ + (p_B - p_{AB})\, max\{(1-q)s, (1-\varsigma(0)) r_R^T + \varsigma(0)\mathrm{r}_B^L\} \\ + p_{AB}\, max\{(1-q)s, (1-\varsigma(0))\mathrm{r}_A^L + \varsigma(0)\mathrm{r}_B^L\} \end{aligned} \quad (4)$$

The outcome of the maximum condition in each state of the world is driven by the deposit insurance. If in a given state of the world the project cannot repay the agreed repayment rate, 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 lender can lose is the difference between what projects should have repaid and the amount that depositors expected to receive. Any shortfall beyond this is picked up by the deposit insurance.

Obviously, $r_S^T < r_R^T$.

In equilibrium, all financial resources of the banks are allocated, i.e. the market clears when

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

and the share of the risky projects of type B, ς, is

$$\varsigma(0) = \frac{\Phi_B^T(0)}{\Phi_A^T(0) + \Phi_B^T(0)}.$$

2.1.2. The emerging stage

First, we consider the scenario in which there are relatively few innovative banks, i.e. we assume that $0 < \mu \leq \mu^*$, for some μ^*.

When innovative banks start entering the market, there will be a small amount of them relative to the proportion of type A projects available in the market. In this case the repayment rates in the marketplace will be driven by competition between traditional banks. Traditional banks cannot distinguish between projects A and B and, therefore, in equilibrium, traditional banks offer all those risky projects that borrow from traditional banks a repayment rate r_R^T. The emerging stage has the following relationship between the repayment rates:

Proposition 1. Assume that there exists $\varepsilon > 0$ such that for all μ in the interval $(0, \varepsilon]$ a traditional bank can repay depositors whenever type A projects succeed. Then there exists $\mu^* \in (0, \varepsilon]$ such that for every $\mu \in (0, \mu^*]$ the following is true:

- *r_S^T is strictly decreasing and r_R^T is strictly increasing in the interval (0, μ^*]*
- $r_R^T = r_R^I$.

To see why Proposition 1 holds it is helpful to note that ς is an increasing function of μ, (since innovative banks are only incentivized to take type A projects, hence the traditional banks have all type B projects and a share of type A projects). As indicated, at μ =0 the repayment rates offered by the traditional banks must satisfy (4). Now, let us consider a marginal increase in μ . There are three potential cases to consider.

Case 1: In the interval $(0, \varepsilon]$ traditional banks remain solvent even if both projects fail simultaneously.

If the banks are able to repay depositors in all states of the world, (4) takes the form:

$$r_S^T = (1 - p_A - p_B + p_{AB})r_R^T + (p_A - p_{AB})\left(\left(1 - \varsigma(\mu)\right)\mathrm{r}_A^L + \varsigma(\mu)r_R^T\right)$$
$$+ (p_B - p_{AB})\left(\left(1 - \varsigma(\mu)\right)r_R^T + \varsigma(\mu)\mathrm{r}_B^L\right) \quad (5)$$
$$+ p_{AB}\left(\left(1 - \varsigma(\mu)\right)\mathrm{r}_A^L + \varsigma(\mu)\mathrm{r}_B^L\right).$$

Differentiating (5) by ς we obtain:

$$\frac{\partial r_S^T}{\partial \varsigma} = (p_A - p_{AB})\,(-\mathrm{r}_A^L + r_R^T) + (p_B - p_{AB})\,(-r_R^T + \mathrm{r}_B^L) + \; p_{AB}\,(-\mathrm{r}_A^L + \mathrm{r}_B^L)$$

or, when rearranged:

$$\frac{\partial r_S^T}{\partial \varsigma} = (p_A - p_B)\,(r_R^T - r_B^L) - \, p_A\,(\mathrm{r}_A^L - \,\mathrm{r}_B^L). \tag{6}$$

Given that $r_R^T > \mathrm{r}_B^L$, and $(1) - (3)$, (6) is always negative.

Case 2. In the interval $(0, \varepsilon]$ the traditional banks are solvent providing at least one of the projects does not fail.

In this case, (4) becomes:

$$r_S^T = (1 - p_A - p_B + p_{AB}) r_R^T + (p_A - p_{AB})\, \left((1 - \varsigma(\mu)) \mathrm{r}_A^L + \varsigma(\mu) r_R^T \right)$$

$$+ (p_B - p_{AB})\, \left((1 - \varsigma(\mu)) r_R^T + \varsigma(\mu) \mathrm{r}_B^L \right) + \, p_{AB}(1 - q)s\,. \tag{7}$$

Differentiating (7) with regard to ς we obtain:

$$\frac{\partial r_S^T}{\partial \varsigma} = (p_A - p_{AB})\,(-\mathrm{r}_A^L + r_R^T) + (p_B - p_{AB})\,(-r_R^T + \mathrm{r}_B^L)$$

or when rearranged:

$$\frac{\partial r_S^T}{\partial \varsigma} = (p_A - p_B)\,(r_R^T - \mathrm{r}_B^L) - (p_A - p_{AB})\,\,(\mathrm{r}_A^L - \mathrm{r}_B^L). \tag{8}$$

Given (1) - (3), $r_R^T \; > \; r_B^L$ and $p_A > p_{AB}$, (8) is negative.

Case 3. In the interval $(0, \varepsilon]$ traditional banks are solvent if projects A succeed and projects B fail but the banks cannot repay depositors if projects B succeed and A fail.

In this case (4) becomes:

$$r_S^T = (1 - p_A - p_B + p_{AB}) r_R^T + \, (p_A - p_{AB})\,(\,(1 - q)s\,) +$$

$$+ (p_B - p_{AB})\,(\,(1 - \varsigma(\mu)) r_R^T + \varsigma(\mu) \mathrm{r}_B^L) + p_{AB}\,(1 - q)s\,. \tag{9}$$

Differentiating (9) with respect to ς gives:

$$\frac{\partial r_S^T}{\partial \varsigma} = (p_B - p_{AB})\,(\mathrm{r}_B^L - r_R^T) \tag{10}$$

Given that $p_B - p_{AB} > 0$ and $\mathrm{r}_B^L - r_R^T < 0,(10)$ is negative.

Hence, regardless of which of the above cases holds, the right-hand side of (4) is a declining function of μ, i.e. the difference between r_S^T and r_R^T must increase. If both repayment rates increase, then the aggregate number of borrowers with risky and safe projects would decline contradicting the market clearing condition. Similarly, if both rates decrease, 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 the traditional banks charge the borrowers with the safe projects falls and the rate charged to the borrowers of the risky projects rises. The innovative banks will charge the highest repayment rate they can subject to finding sufficient number of borrowers with type *A* projects. When there are few innovative banks, then this rate will be equal to the rate the traditional banks charge their borrowers with risky projects.

The relationship between r_S^T and r_R^T will be given by one of the three cases above as μ initially increases above 0. However, as the size of the innovative sector increases, the proportion of type *A* projects in a traditional bank's portfolio of risky projects declines. Unless r_B^L is greater than or equal to $(1-q)s$, there must come a point when the proportion of type *A* projects is insufficient to offset the type *B* project's shortfall between $(1-q)s$ and r_B^L when they fail. At this point, the traditional banks that specialize in risky projects will be unable to repay their depositors whenever type *B* project B fail. If r_B^L is strictly less than $(1-q)s$, we define μ^* as the minimum μ where this happens (μ^*will obviously depend on the specific parameter values). If r_B^L is greater than or equal to $(1-q)s$, μ^* is defined as the minimum μ where all type *A* projects are served by the innovative banks and $r_A^I = r_R^T$.

Subject to the assumption that when $\mu = 0$, the traditional banks will be able to repay depositors whenever type *A* projects are a success, the introduction of innovative banks raises the repayment rates to all risky projects at both the traditional and the innovative banks, and reduces the repayment rates for the safe projects.

2.1.3. The non-segregated intermediate stage

Initially, consider the case where r_B^L is less than $(1-q)s$. In other words, for μ greater than μ^*, the traditional banks cannot repay depositors when type *A* projects are successful and type

B projects fail. If at this point, the traditional banks cannot repay depositors when type *B* projects are successful and type *A* projects fail, then (4) becomes:

$$r_S^T = (1 - p_A - p_B + p_{AB})r_R^T + (p_A - p_{AB})(1-q)s + \\ + (p_B - p_{AB})(1-q)s + p_{AB}(1-q)s \quad (11)$$

In this case an increase in μ has no impact on the relative repayment rates of the safe and the risky projects, which remain constant for an interval above μ^*. However, as the innovative sector grows, it must eventually be the case that the portfolios of risky projects of the traditional banks will be able to cover the repayment to depositors. In this case (4) becomes:

$$r_S^T = (1 - p_A - p_B + p_{AB})r_R^T + (p_A - p_{AB})\left((1-\varsigma(\mu))\mathrm{r}_A^L + \varsigma(\mu)r_R^T\right) + \\ + (p_B - p_{AB})(1-q)s + p_{AB}(1-q)s. \quad (12)$$

The derivative of (12) with respect to ς is:

$$\frac{\partial r_S^T}{\partial \varsigma} = (p_A - p_{AB})(r_R^T - \mathrm{r}_A^L) > 0.$$

Small increases in the innovative sector at this point will reduce the repayment rates charged to all the depositors with the risky projects and increase the repayment rates of the borrowers with safe projects.

As the innovative sector grows (μ increases above μ^*), there will be a point when, at the prevailing equilibrium prices, all borrowers with type *A* projects will be served by the innovative banks only. Let μ^{**} be the lowest μ where this happens.

This indicates the following Proposition:

Proposition 2. If r_B^L *is less than* $(1-q)s$*, there exists an interval* $\mu^* < \mu < \mu^{**}$ *where*

- r_S^T *is a weakly increasing function of* μ
- r_R^T *is a weakly decreasing function of* μ

Note, that if $r_B^L \geq (1-q)s$, then $\mu^* = \mu^{**}$.

2.1.4. The segregated intermediate stage

If $\mu > \mu^{**}$, the innovative sector has grown enough to absorb all borrowers with type A projects at the prevailing prices (in particular, when $r_R^T = r_R^I$). As μ increases slightly above μ^{**}, innovative banks compete to lend to the borrowers with type A projects. The competition will reduce the repayment rates offered by the innovative banks and the price will drop to the point when enough borrowers with type A projects are brought into the market to meet the demand by the innovative banks. For $\mu > \mu^{**}$ the traditional banks will know that their borrowers with risky projects are all type B and will set the repayment rates to reflect this. For notational convenience we denote r_R^T by r_B^T whenever $\mu > \mu^{**}$.

Proposition 3. There exists μ^{***} *such that for every* $< \mu \in (\mu^{**}, \mu^{***}]$ *the following is true:*

- r_S^T *is an increasing function of* μ
- r_A^I *is a decreasing function of* μ
- r_B^T *is an increasing function of* μ

Note, for $\mu \in (\mu^{**}, \mu^{**} + \varepsilon)$, in equilibrium, the borrowers with the safe projects borrow at the repayment rate r_S^T, borrowers with type A projects borrow at the repayment rate r_A^I, and borrowers with type B projects borrow at the repayment rate r_B^T.

The overall market clearing condition is:

$$\Phi_S(r_S^T) + \Phi_A(r_A^I) + \Phi_B(r_R^T) = D. \tag{13}$$

Moreover,

$$\Phi_S(r_S^T) + \Phi_B(r_R^T) = (1-\mu)D. \tag{14}$$

Putting (13) and (14) together gives

$$\mu D = \Phi_A(r_A^I). \tag{15}$$

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

(14) also implies that when μ increases $\Phi_S(r_S^T) + \Phi_B(r_R^T)$ decreases. Moreover, given that

$$r_S^T = r_R^T - p_B(r_R^T - max\{\mathrm{r}_B^L, (1-q)s\}), \quad (16)$$

we have

$$\frac{\partial r_S^T}{\partial r_R^T} = (1 - p_B) > 0.$$

This means that r_S^T and r_R^T move in the same direction. If they both declined, $\Phi_S(r_S^T) + \Phi_B(r_R^T)$ would increase. However, we have already shown that $\Phi_S(r_S^T) + \Phi_B(r_R^T)$ decreases. This means that r_S^T and r_R^T must be increasing functions of μ.

Note, that as μ grows competition amongst the innovative banks decreases r_A^I until

$$r_S^T = r_A^I - p_A(r_A^I - max\{\mathrm{r}_A^L, (1-q)s\}). \quad (17)$$

Thus, the critical value μ^{***} is the μ that solves the system of equations (13), (16) and (17).

2.1.5. The mature stage

By the time the market share of the innovative banks reaches μ^{***}, it is not optimal for the innovative banks to lower the repayment rates any further. Thus, when $\mu > \mu^{***}$, the innovative sector has no incentive to grow any further. This leads to:

Proposition 4. When $\mu > \mu^{***}$, the following holds

- r_S^T *is a constant function of* μ
- r_A^F *is a constant function of* μ
- r_B^T *is a constant function of* μ

Putting all these propositions together gives a set of repeyment rates as shown in Figure 1.

********** insert Figure 1 here ************

Note that Proposition 1 was contingent on the assumption that there exists an $\varepsilon > 0$ such that for all $\mu \in (0, \varepsilon]$ the traditional banks can repay depositors whenever type A projects succeed. In reference to Figure 1, we should point out that if such an interval does not exist, then $\mu^* = 0$. As already noted, if r_B^L is greater than or equal to $(1-q)s$ then $\mu^* = \mu^{**}$.

3. Implications and extensions

In this paper we explore a scenario where traditional banks compete with innovative (fintech) banks that have ML and similar technologies to better 'diagnose' the quality of borrowers. The paper shows that there are significant non-linearities associated with the entry of innovative banks. We identify four phases. However, the core phases are the emerging phase and the segregated intermediate phase. In the emerging phase the dominant form of competition is that between traditional banks. By distorting the mix of risky projects within traditional banks, the entry by innovative banks creates a negative externality, driving up the repayment rates for all risky borrowers and generating a 'flight to safety'. In the segregated intermediate phase competition by innovative banks for the lower risk projects starts to bite and impact on the repayment rates for borrowers. The desire to suck more lower risk projects into the market begins to drive down borrowing rates for the lower risk borrowers and raise rates for the higher risk borrowers, which begins to unravel the negative externality. In this section, we close the paper with comments on the model and some simple extensions.

3.1. Competition, stability and concentrated business models.

We begin by addressing the relationship between competition, concentrated business models and stability, an issue that has attracted considerable attention in the academic and policy literature. 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 what happens if the greater competition arises on the asset side, which is the case discussed in this paper. Greater competition for borrowers tends to reduce repayment rates. This is likely to reduce moral hazard incentives for borrowers, hence reduce the probability of borrower default (see Boyd and De Nicolo, 2005). In terms of stability of banks, as opposed to projects, the position is less clear. It is true that increasing competition reduces the probability of individual borrowers' default. However, it is also likely to reduce the profit to the 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 (see Martinez-Miera and Repullo, 2010). Both Boyd and De Nicolo (2005) and Martinez-Miera and Repullo (2010) have competition for borrowers from symmetrically positioned banks. In this paper, we analyse the impact of asymmetric forms of competition. The impact of competition from innovative banks on the stability of the banking system depends on the circumstances, but the model gives insight as to which factors matter and when.

Instability in our model can arise from two core sources. One is that entry by innovative banks impacts on the allocation of projects across banks. Relative to the share in the economy, innovative banks have disproportionately more of the better projects whilst traditional banks have less of them. Call this the allocation effect. The other is that competition affects the relative repayment rates, which we call the relative rate effect.

Considering the relative rate effect, other things being equal, a lower repayment rate for safe projects in conjunction with a higher repayment rate for risky projects will encourage more safe projects and lower risk ones, and vice versa. It is clear from Figure 1 that in the emerging stage of development, that the relative rate effect can generate a flight to safety. That is, falling repayment rates for safe projects and rising rates for risky projects, encourage more safe projects and reduce risky projects, even though innovative banks can identify the lower risk projects. Thus, the relative rate effect associated with more innovative bank competition tends to reduce aggregate risk taking by banks in the emerging phase of development. It follows, if the parameters values are such that at least one of the risky project types fail to cover repayments to depositors when the project fails, that a change in relative repayment rates induced by an increase in the innovative sector, could be associated with greater financial stability in the banking sector. In general, the relative rate effect tends to increase financial stability in the emerging stage, but in the intermediate stage tends to reduce it.

The allocation effect, however, can offset the impact of the relative rate effect. An interesting example arises when repayments to the bank from type A projects when they fail are greater than the repayments to depositors 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. In the face of small increases in innovative banks, traditional banks will still find that their portfolio of risky projects will, in aggregate, be able to repay depositors, so the relative rate effect will dominate. Call the traditional banks with both risky and safe borrowers general traditional banks. However, at some point during the emerging stage, there will be enough innovative banks that the return from their mix of the risky projects in the traditional banks no longer covers the repayments to depositors if both projects fail. Note that, the general traditional banks will still be sound because there will be enough profit from the safe projects to make up the shortfall. However, the traditional banks have an alternative strategy. They could become banks with concentrated business models, either concentrating on risky projects or concentrating on safe projects. Once the mix of risky projects in the general tradition banks is no longer able to meet the required return to depositors, then the general traditional banks will be unable to compete against the traditional banks with concentrated business models. The reason for this is deposit insurance. Any proportion of the innovative banks beyond this point is associated with concentrated business models for all the banks. Hence, there is a critical proportion of innovative banks such that risky traditional banks dominate the traditional banking market from this point onwards, and thus some risky traditional banks will fail in some states of the world. This generates instability in the banking system, even though a general traditional bank would be solvent should they exist. Note, that this problem will happen whenever there is an emerging stage of development and the higher risk projects cannot cover repayments to depositors when they fail.

3.2. The market for big data and the cost of innovative banking

In this sub-section we discuss what happens in the model if innovative banks have to pay a cost to separate projects, and discuss the associated issue of what happens in the realistic scenario that some of this cost arises because of the need to obtain access to big data that is held by other parties.

Throughout the emerging and the intermediate stages of development, borrowers with type A project pay a higher repayment rate than they would have paid if all types of project were observable by all banks. It follows that if innovative banks have no greater costs than the traditional banks, they then earn a greater return than the traditional banks. Consequently, if there is a cost (per unit of loan) to becoming an innovative bank, essentially a cost to be able to separate project types, then the model would look very similar providing this cost is small. The main difference is that there would be no mature phase. The reason being that there would be no incentive for the innovative banks to continue to enter up to the point where repayment rates are equal to what they would be with full information. There needs to be a wedge between the repayment rates for type A projects with the innovative banks and full information to cover the cost of the innovative banking. However, the core model would be very similar.

Things become different, however, if the costs of innovation are not small. One possibility is that all innovative banks do not have the same cost. If some innovative banks have very low costs but costs differ sufficiently across innovative banks there may be a ceiling on the proportion of innovative banks that can exist in the market. The ceiling could be in any of the first three stages. Another possibility arises if costs are not too dissimilar across banks but even the lowest cost is so high that entry for the innovative banks is not profitable at the traditional banks repayment rate (r_R^T) when there are no innovative banks. Then there is no incentive for the innovative banks to enter unless the share of the market by the innovative banks is above some critical number. Essentially, we have a network effect since the repayment rate an innovative bank can charge is increasing in the number of the innovative banks in the market. There is a positive minimum proportion of innovative banks that are needed for the innovative banks to survive. Similarly, there will be a maximum share beyond which the innovative banks cannot make a profit. Thus, there can be multiple equilibria. We return to this issue in the following subsection.

An interesting special 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 projects of type A from projects of type B. 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 GAFAs (Google, Amazon, Facebook and Apple). A simple case arises if 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 maximises the aggregate profit (before subtracting access fees) of the innovative banks. Thus, the owner of the big data would set a fee that limits the size of the innovative market. Unlike a normal monopoly, there is a region, the emerging phase, when the price that can be charged for access to the big data is increasing as the proportion of innovative banks purchasing the data increases, so there is no tradeoff between price and quantity in the emerging stage. However, the long run equilibrium fee would trigger the problem of a network effect as identified above. So, if the maximum size of the innovative sector is growing continuously over time then the initial access fee needs to be low enough to avoid a multi-equilibrium problem. Indeed, what we would see is a low price as innovative banks emerge with the fee rising as the sector grows. It is possible that this price would need to be negative if a network problem exists even in the absence of access fees.

The maximum potential access fee arises when the innovative banks have entered in sufficient numbers to maximise the repayment rates charged by the innovative banks. However, the innovative sector will probably grow beyond the emerging stage. For example, suppose that traditional banks can only survive if both projects succeed at the upper bound of the emerging stage. At this point the rates charged to risky borrowers by traditional banks, and innovative banks, will be constant for a small increase in the innovative sector. The maximum access fee is constant but volume increases, hence in this case the data owner will gain by an increase in the size of the innovative sector beyond the emerging stage. Eventually the maximum charge that can be made to each innovative bank will be falling as the proportion of innovative banks increases, hence there is the usual volume-price tradeoff for the supplier of big data.

3.3. Additional policy implications

Section 4.1 had already discussed the relationship between innovative banking and stability, and here we will consider some other broader issues.

The first point of interest concerns the welfare effects of innovative banking in the model. Obviously, the welfare effects of changes in the proportion of innovative banking 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 emerging phase every increase in the proportion of innovative banks leads to a drop in welfare. The most obvious case arises if the traditional banks are solvent even when all projects fail. At this point, there is no call on deposit insurance arising from a small increase in the share of the innovative banks. However, the prices that are charged to risky projects are not based on the true mix of low and high-risk projects but as if there are too many high-risk projects. Hence, the tradeoff 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 emerging phase. However, once one gets to the intermediate stage, the price of the lower risk projects falls and the problems begin to unwind, suggesting at this point welfare is likely to increase with increases in the innovative sector.

At μ^{***}, both A and B type projects are required to pay the rates that would arise if there were full information. However, whether welfare is higher at $\mu = \mu^{***}$ or without any innovative sector depends on two counter effects. Without any innovative sector, both types of risky projects pay the same rates, so there are too many high-risk projects relative to low-risk projects. It is possible at this point that the 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 risky projects, there is the right mix of risky and safe projects. On the other hand, when $\mu = \mu^{***}$, we have the right allocation between types of risky project but (assuming $r_B^L < (1-q)$) we have a moral hazard problem induced by the deposit insurance (i.e. banks choose too many risky projects). Whether welfare is higher with no innovative sector or if $\mu = \mu^{***}$ depends on whether the moral hazard causes more welfare loss than the misallocation due to the inability to separate As from Bs. If welfare is greater at μ^{***}, 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 sector (or to be more precise, discourage those fintech banks whose only benefit is the ability to separate between project types). We return to this point after a brief discussion of capital requirements and stability.

In the model, we assume that the opportunity of cost of capital, c, is greater than the cost of deposits.[7] As a result, capital is relatively expensive and banks will choose no more capital than

[7] 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.

the minimum ratio, q, set by the prudential regulator. Thus, q is a policy tool that the regulator can use to achieve various objectives. Changing q typically improves stability. For example, in Section 4.1, we discussed how the allocative effect could lead to a shift to concentrated business models and, hence, a reduction in stability at an interior point of the emerging phase. Holding other parameter values constant, an increase in q would delay the point when traditional banks must adopt concentrated business models to survive and, hence, increases the proportion of the emerging phase where 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 lead to lower rates for safe projects. Thus, there are more safe projects and more stability.

Returning to the welfare discussion, one way that innovative banks could be encouraged would be to have a lower capital requirement for innovative banks compared to traditional banks. This would seem consistent with projects of type A being lower risk, hence generating less moral hazard concerns. However, under current international prudential policy (Basel III), the exact opposite is the case, i.e., innovative banks are typically required to hold far higher capital ratios than traditional banks. The reason for this is that traditional banks are far more likely to have internal risk based (IRB) models whereas new banks, having limited lending history, will typically have to follow the Standardised Approach (SA), which attaches standardised risk weights to loans instead of bespoke risk weights. Since the SA weights cover all banks without IRB models regardless of type, the SA weights tend to be significantly higher than IRB risk weights.[8] Policies to change this imbalance would clearly help the innovative sector.

References

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

Berg, T., Burg, V., Gombović, A., Puri, M. 2019. On the rise of Fintechs: Credit scoring using digital footprints. Review of Financial Studies.

Boyd, J. H., De Nicolo, G. 2005. The Theory of bank risk taking and competition revisited. Journal of Finance 60: 1329–1343.

[8] The IRB weights are not always in the public domain, but those that are available show significant differences between IRB and SA. For example, in the UK in 2015 the average risk weight for a mortgage with LTV<50% from a large lender was 7.9% compared to the SA risk weight of 35%.

Claessens, S., Frost, J., Turner, G., Zhu, F. 2018. Fintech credit markets around the world: size, drivers and policy issues. BIS Quarterly Review, available at https://www.bis.org/publ/qtrpdf/r_qt1809e.pdf .

Crawford, G.S., Pavanini, N., Schivardi, F. 2018. Asymmetric information and imperfect competition in lending markets. 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.

Dorfleitner, G, Priberny, C., Shuster, S., Stoiber J., Weber M., 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 & Finance 64: 168-187.

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., Shin H. S., Zbinden P. 2019. BigTech and the changing structure of financial intermediation, Economic Policy, forthcoming (also published as BIS Working Papers, 779)

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

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

Iyer, R., Khwaja, A. I., Luttmer, E. F., Shue, K. 2015. Screening peers softly: Inferring the quality of small borrowers. Management Science 62: 1554-1577.

Martinez-Miera, D., Repullo, R. 2010. Does competition reduce the risk of bank failure? Review of Financial Studies 23: 3638-3664.

OECD (2020) Financing SMEs and Entrepreneurs 2020 https://www.oecd.org/cfe/smes/financing-smes-and-entrepreneurs-23065265.htm

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: 1–15.29.

Figure 1. Repayment rates of risky and safe borrowers as a function of the market share of the innovative banking sector (μ).