



**UNIVERSITY
OF WARWICK**

Economics

Bank Lending Practices and Size-Dependent Credit Treatment: An Endogenous Audit Threshold

Ghasan Asbool and Damiano Turchet

October 2026

No: 1636

ISSN 2059-4283 (online)
ISSN 0083-7350 (print)

Warwick Economics Research Papers

Bank Lending Practices and Size-Dependent Credit Treatment: An Endogenous Audit Threshold*

Ghasan Asbool[†]

Damiano Turchet[‡]

Abstract

How can credit-market frictions generate heterogeneous investment responses across firms of different sizes? We develop a model in which lenders choose whether to incur auditing costs when lending to firms whose idiosyncratic productivity shocks are otherwise unobservable. The lender's optimisation problem gives rise to an endogenous capital threshold that separates firms into different credit-treatment regimes. As aggregate conditions change, movements in this threshold alter firms' access to credit and generate non-uniform credit responses across the size distribution. The model provides a financial-frictions channel through which medium-sized firms may experience disproportionately larger investment declines during recessions, through a credit channel. We thereby offer a foundation for studying the investment patterns, such as the ones exhibited by UK firms during the financial crisis and by German firms over 1970–1986.

Keywords: *Bank Lending Practices, SME, Corporate Banking, Credit Discrimination, Investment Specific Technology Shocks*

JEL Classification: E44, G21

*Preliminary work. We thank participants to the seminars at the University of Warwick, the 2026 Warsaw International Economic Meeting and the 3rd Modern Finance Conference, Krakow, Poland for helpful comments and suggestions. All errors are our own.

[†]Department of Economics, University of Warwick, Coventry CV4 7AL, UK. E-mail: G.Asbool@warwick.ac.uk

[‡]Department of Economics, University of Warwick, Coventry CV4 7AL, UK. E-mail: Damiano.Turchet@warwick.ac.uk

1 Introduction

Most of the literature discussing the variation in performance between firms based on their size has focused on dividing firms into two groups, Small and Medium Enterprises (SMEs) as a single block and large firms. Little work, at least within economics, has tried to discover theoretically the variation within SMEs. It is important to investigate this variation, since more than 95 per cent of firms around the world belong to this category. More than half of the workers in emerging economies are employed by firms with fewer than 100 employees (Ayyagari et al., 2011). SMEs are equally important in developed economies. For instance, according to the UK's Department for Business and Trade (DBT, 2025), SMEs accounted for 99.9 per cent of all private sector businesses, 60 per cent of private sector employment and 51 per cent of private sector turnover at the start of 2025.¹ The importance of SMEs extends beyond the absolute numbers and proportions they account for. Evidence of a negative relationship between firm size and net growth rates was found using US firms' employment dynamics datasets in Haltiwanger et al. (2010) and Neumark et al. (2011).

Many studies reflect this pattern of difference in performance between small and large firms, especially during recessions and the following recoveries, with small firms suffering proportionately more and exhibiting a slower recovery (see Sahin et al., 2011; Fort et al., 2013). The main explanation provided for this pattern is the severe reduction in credit extended by banks to small firms. Those firms are typically more reliant on bank lending, and hence are more affected by banking distress. The greater financial limitations faced by SMEs as opposed to large firms have translated into different real activity. Using US data, Duygan-Bump et al. (2015) find that small firms' financial limitations played an important role in contributing to the high unemployment during the recession that ensued from the 2008 financial crisis, where small firms' workers were more likely to become unemployed. Using UK data, Benito et al. (2010) show more generally that the contraction in credit supply during the financial crisis caused a sharp fall in business investment.

Rather than focusing on the aggregate supply of credit, this paper examines how banks differentiate their lending practices across firms. In many commercial banks, lending activities to firms are split into two main divisions, each dealing with firms of a certain size category. Usually a firm's turnover or net worth is used as a proxy by banks to determine in which division the firm will be dealt with. The two divisions differ not only in the resources they allocate but also in the evaluation methods they apply: corporate banking divisions that deal with large firms dedicate considerable time and better qualified staff to assessing the quality of a borrower's projects and financials, whereas SME banking relies more heavily on rigid credit scoring and on the traditional capital gearing approach, asking for high collateral against the

¹There is no universally agreed standard on classifying firms based on their size or on other criteria such as their revenue or net worth. Definitions differ across countries and sometimes even between institutions within the same country. Among the standards used for classifying firms are workforce size, assets and turnover (Beck, 2013).

credit provided (Albareto et al., 2011; Binks and Ennew, 1997). Accordingly, when loans are approved, large firms enjoy more flexible and less collateral-intensive terms. The conventional explanation centres on the high costs of conducting thorough credit assessments, which, relative to the smaller loan amounts typically demanded by SMEs, render such assessments economically unviable for lenders.

Motivated by this conventional explanation, we investigate whether size-dependent credit treatment can arise endogenously in a model in which firms differ only in their initial capital endowments. To the best of our knowledge, existing theoretical models that generate size-dependent credit treatment and loan contracts rely on exogenously imposed heterogeneity between small and large firms. Crucially, this heterogeneity extends to at least one dimension other than firms' initial size. Firms are assumed to be either of multiple types, differing in their ability to repay (Bester, 1985; Besanko and Thakor, 1987), in the productivity of their technology (Becsi et al., 2005), or in the investment projects they can access (Caminal and Matutes, 1997; Villas-Boas and Schmidt-Mohr, 1999). Alternatively, some models abstract from imposing such differences and simply assume the heterogeneous credit treatment exogenously (Bernanke et al., 1999). Here the only difference between borrowers is the capital they are endowed with.

The mechanism through which these differences give rise to size-dependent credit treatment depends on uncertainty over production outcomes. Production by the borrower is subject to an idiosyncratic shock whose realisation is not observable before production takes place. The lender can verify the shock's realisation through an audit, but doing so incurs a positive fixed cost. If the lender audits and the project is found to be successful, he lends without imposing collateral requirements, since the borrower can repay out of output. If he does not audit, he still lends, but he fully secures the loan against the borrower's undepreciated capital stock. Thus, the borrower may be unable to obtain her optimal loan. Auditing therefore enables the lender to extend a larger loan, which results in an increase in surplus, therefore, obtains a greater payoff albeit at a fixed audit cost. As the borrower's capital increases, the difference between the loan amounts extended with and without auditing widens, thereby increasing the lender's incremental payoff from auditing. Since the audit cost remains fixed, there exists a threshold level of capital $\hat{k}$ at which this incremental payoff exactly offsets the cost of auditing. Accordingly, auditing is optimal for the lender if and only if $k \geq \hat{k}$, which makes borrowers on either side of the threshold face different credit treatments despite being identical in all other dimensions other than the amount of capital they own.

A second empirical observation motivating the model concerns the non-monotonic relationship between firm size and investment under credit constraints. Using UK firm-level data covering 1997–2009, Crawford et al. (2013) find that investment declined most sharply relative to its pre-crisis linear trend among medium-sized firms, followed by small and then large firms. Audretsch and Elston (2002), using German data, report a qualitatively similar pattern. This empirical pattern is at odds with the conventional intuition in the financial frictions literature,

according to which investment should decline most sharply among smaller firms, which tend to face tighter credit constraints. Our model offers a possible explanation for this pattern. Adverse shocks shift the endogenous auditing threshold $\hat{k}$, disproportionately affecting borrowers near the threshold. Some of these borrowers consequently switch from audited to unaudited lending, resulting in smaller loans than would otherwise have been extended. To examine the implications for investment, however, the model will need to be extended. In its present form, the loan finances working capital, while the borrower's capital stock is fixed within the period. The analysis therefore focuses on the endogenous auditing threshold and its properties.

Section 2 reviews the literature most relevant to the theoretical framework and provides the background motivating the model. Section 3 introduces the model, while Section 4 characterises the equilibrium and derives the endogenous auditing threshold. Section 5 establishes the conditions under which the threshold exists and examines its comparative statics. Finally, Section 6 discusses the limitations of the current framework and outlines possible extensions, including the introduction of firms' investment decisions.

2 Literature Review

This section reviews three distinct strands of literature relevant to the analysis: bank lending practices, credit discrimination and financial frictions in macroeconomics, and investment-specific technology shocks. Together, these provide the theoretical and empirical background for the model and its proposed extensions.

2.1 Bank Lending Practices

In much of the theoretical literature on lender–borrower relationships, banks are modelled as single entities, with little attention paid to their internal organisational structure. In practice, however, commercial banks commonly organise their lending activities into separate divisions, with firms assigned to these divisions according to their size. A firm's turnover or net worth is typically used as a measure of size in determining which division handles its lending. This segmentation is reflected in differences in resource allocation, lending products, service delivery and borrower assessment procedures (Beck, 2013). Although banks may operate more than two lending divisions, we restrict our analysis to two, as this is sufficient to capture the size-dependent differentiation in lending practices that is central to our model. Hereafter, we refer to these two divisions as SME banking and corporate banking. The former serves firms below a certain size threshold, while the latter serves larger firms.

In presenting the characteristics of the two divisions we make extensive use of Albareto et al. (2011), who conducted a survey on behalf of the Bank of Italy on banks conducting lending activities to firms. Their sample covers 82.2 per cent of outstanding loans made by Italian banks to non-financial firms in 2007. In assessing the creditworthiness of their business customers,

banks rely on both quantitative techniques and soft information. However, the relative importance of each differs across lending divisions according to the size of the firms they serve. In corporate banking divisions, banks tend to employ more qualified staff and dedicate more time and effort to assessing the quality of large firms' projects and financial positions. Rigid credit scoring techniques and statistical models are given less relative importance, while greater emphasis is placed on qualitative information, such as the borrower's organisational structure, management credentials, the quality of its investment projects and the extent of its banking relationships with competing banks. This allows for greater reliance on judgement in lending decisions. When approved, loans to large firms typically require less collateral relative to the loan amount than those extended to smaller firms. Several factors help explain these lower collateral requirements. Large firms tend to have greater bargaining power, as they maintain a wider network of banking relationships and can more easily switch between lenders. They may also have access to commercial paper markets, which remain largely inaccessible to smaller firms. Furthermore, the more thorough credit assessments conducted by banks can help identify higher-quality projects and reduce the risk of default on approved loans (Albareto et al., 2011; Cressy and Olofsson, 1997).

This is consistent with some of the predictions of relationship lending theories, although these theories do not explicitly examine the segmentation of lending activities within banks. They are based on the existence of asymmetric information between borrowers and lenders, which can give rise to differences in credit treatment. Two main predictions follow. First, the cost to borrowers of switching lenders declines as borrower transparency improves. Larger firms tend to be more transparent, as they generally have better financial records and more established reputation. Second, the cost of asymmetric information declines with borrower size, as larger loans have lower monitoring costs per unit lent (Chodorow-Reich, 2014). The literature also examines how banks can address asymmetric information through borrower screening. Bester (1985) and Villas-Boas and Schmidt-Mohr (1999) develop models in which screening can substitute for credit rationing.

In SME banking, banks devote less time and fewer resources to assessing the quality of borrowers' projects. The number of potential SME customers is large, while the profit earned from each banking relationship is typically lower than in corporate banking. Technological progress has also reduced the cost of processing data, leading banks to rely more heavily on credit scoring models and statistical techniques when assessing SME borrowers. This has allowed banks to allocate more resources to screening and auditing larger firms (Albareto et al., 2011). The declining dependence on soft information is also a consequence of how much information is available in the first place, since its availability and quality are strongly linked to firm size. Using data from the American National Survey of Small Business Finances, Petersen and Rajan (1994) find that small firms are more likely to face credit rationing. This is because the lack of sufficient information means that lenders are less able to screen out potential defaulters. Cowan

and Cowan (2006) find, in a survey of US banks, that the lesser reliance on soft information for SME customers meant that banks depended more on traditional creditworthiness criteria such as the availability of collateral. Credit rationing arising from a lack of information has been addressed theoretically. For example, Jaffee and Russell (1976) build a model in which lenders ration credit when they cannot distinguish the quality of borrowers. A similar outcome arises in the model we present in this paper, where some borrowers are unable to obtain their optimal borrowing demand because of uncertainty about the quality of their projects.

Binks and Ennew (1997), using data on six thousand UK small businesses, study the practices banks follow in dealing with their SME customers. Banks combine income and capital gearing approaches in evaluating loan applications, though the balance between them varies with firm size. The income approach evaluates loan applications on the basis of a firm's cash flow relative to the loan amount, rather than its collateral relative to loan size. This provides a more direct assessment of the borrower's ability to repay the loan. However, it requires detailed knowledge of the firm's operations and the market in which it operates, as well as access to relevant information provided by the borrower. This is less likely in the case of SMEs. It is also the case that such evaluations are costly for banks to carry out and, given the relatively small size of SME loans, are often unprofitable to undertake. Banks may also have less incentive to offer flexible lending terms to SMEs, given their relatively low rates of switching between banking providers. According to the SME Finance Monitor, only 4 per cent of UK SMEs had changed their main bank over the preceding three years (BVA BDRC, 2026). Banks therefore rely on the capital gearing method, asking for high collateral against the credit provided. The model that we present formalises this distinction by assuming that when the lender does not audit, the loan must be fully collateralised, whereas auditing allows the lender to extend loans without the same collateral requirement.

Although bank lending practices play a central role in what follows, we adopt a reduced form approach that abstracts from modelling the financial intermediation process. Lenders and borrowers interact directly, with the lenders performing the credit screening that is typically carried out by the banking sector.

2.2 Credit Discrimination and Financial Frictions in Macroeconomics

The large macroeconomic literature on financial frictions deals with informational frictions between lenders and borrowers. Such frictions result in problems of adverse selection or moral hazard, which in turn lead to feedback and propagation channels that amplify initial real or monetary shocks. Bernanke et al. (1996, 1999) name the changes in credit market conditions caused by informational frictions the financial accelerator. Models in this literature differ greatly in their predictions and ingredients, reflecting disagreement about which informational frictions are of paramount relevance. However, they divide broadly between two propagation channels; the net worth (balance sheet) channel and the bank equity (bank lending) channel.

Models employing the net worth channel are more established and address the demand side of credit markets. They depend on high state verification costs, incurred by the lender in order to verify the state of the borrower. Lenders are accordingly discouraged from performing the required audits and instead either ration credit or require firms to pledge collateral against the credit extended to them. The limited wealth of firms creates contracting problems that propagate adverse shocks throughout the economy (Bernanke and Gertler, 1989; Carlstrom and Fuerst, 1997; Bernanke et al., 1999). Kiyotaki and Moore (1997) build a dynamic model in which borrowers' net worth represents a binding constraint on their ability to obtain credit, and in which procyclical movements in asset values affect the collateral they can pledge and hence their ability to borrow and invest. Hart and Moore (1994) develop a model in which borrowers use their assets both as factors of production and as collateral. On the other hand, models employing the bank equity channel address the supply side and are less established. For instance, Den Haan et al. (2003) develop a dynamic equilibrium model in which households' loanable funds are channelled to borrowers through intermediaries via long term relationships. In their model, a sufficiently large adverse shock can destroy enough relationships to collapse the credit market.

Our model follows the net-worth approach to credit frictions, with costly state verification as its only friction. When the lender chooses not to audit, the loan must instead be fully secured against the borrower's undepreciated capital stock. It departs from that tradition in one respect. As noted in Section 1, to the best of our knowledge all the theoretical models in this literature generate discriminating credit treatments by imposing exogenous heterogeneity between small and large firms in one or more dimensions other than their size. In contrast, our model assumes that borrowers differ only in their initial capital endowments. The lender's optimal decision of whether to audit gives rise to an endogenous capital threshold, which determines the credit treatment borrowers receive in equilibrium.

2.3 The Role of Investment Specific Shocks

Investment specific technology (IST) shocks play a role in the model presented in the following section. The rate at which the final good is converted into capital determines both the value of the borrower's outside option and the tightness of her collateral constraint. Unlike neutral technology shocks, which affect the production of all goods homogeneously, investment shocks affect only the production of investment and capital goods (Fisher, 2006). The literature on investment shocks distinguishes between investment-specific technology (IST) shocks and marginal efficiency of investment (MEI) shocks. IST shocks affect the efficiency of producing investment goods relative to consumption goods, while MEI shocks affect the efficiency with which investment goods are transformed into productive capital (Justiniano et al., 2011). We consider IST shocks and implicitly assume that there are no shocks to the marginal efficiency of investment.

The theoretical literature argues that IST shocks represent an important alternative to neutral productivity shocks as a source of business cycles in general equilibrium, with a key role in explaining short run fluctuations as well as long-term movements in indicators such as labour productivity (Greenwood et al., 1988, 1997; Fisher, 1997). Greenwood et al. (2000) calibrate a model using US data on the relative price of new equipment and find that IST shocks account for some 30 per cent of the fluctuations in output, which is substantial given that investment in new equipment represents only 7 per cent of US GNP. Smets and Wouters (2007) and Justiniano et al. (2010) estimate extensions of the New Neoclassical Synthesis DSGE framework using Bayesian methods. Both studies examine the contribution of investment shocks to fluctuations in hours worked and output, but differ in their findings. Justiniano et al. (2010) find that shocks to the marginal efficiency of investment explain a substantially larger proportion of business cycle fluctuations than in Smets and Wouters (2007). Their estimates are also broadly comparable to those obtained by Fisher (2006) using a structural vector autoregression (SVAR). Fisher (2006) adopts the competitive growth model of Greenwood et al. (1997), expressed as a planner's problem and simplified to a single capital good, and conducts a SVAR estimation using the theoretical model to derive the identifying assumptions, finding that IST shocks account for most of the variation in short run labour hours and output. In a competitive economy, the IST level is equivalent to the inverse of the price of investment relative to the consumption good, and it is in this form that the shock enters the model below.

3 The Model

3.1 The Model's Environment

We consider a lender-borrower economy with three goods: two intermediate goods, a generic production factor², X , and capital, k , and the third is a final good, c , which may be used for consumption and for contracting and whose price is normalised to one. The final good can be transferred into capital at a rate q , where q is a technology parameter; $1/q$ is therefore the number of units of the final good that must be exchanged for a unit of the capital good. In a competitive economy, this is the real price of capital (Fisher, 2006).

There is a unit mass of identical households, indexed by n , whom we also call the lenders, and a unit mass of entrepreneurs, indexed by j , whom we also call the borrowers. Since there is a unit mass of each, we assume a one to one matching to take place at the start of the economy³.

Each lender (he) is endowed with a constant amount $\bar{x}$ of the production factor X . He maximises $u(c) = c$ and is therefore risk neutral, and he owns a production technology $g(X)$ which depends only on X and yields one for one in terms of the final good. He may also lend

² X can be thought of as a raw material or as working capital that enters the production of the final good. The endowment is sufficiently large that the lender's resource constraint in lending does not bind.

³This formulation allows us both to abstract from the search problem and let us drop the agents' indices and focus on the matched agents problem.

part of his endowment to the borrower he is matched with. The factor X fully depletes in production, regardless of the technology in which it is used.

Each borrower (she) also maximises $u(c) = c$ and is risk neutral. The assumption that both agents are risk neutral simplifies the analysis and keeps the equilibrium tractable. Borrowers are endowed with physical capital k_j , which is fully observable, but their endowments differ at the start of the economy, so that the aggregate capital stock is $K = \int_0^1 k_j dj$.

Each borrower owns two production technologies, both producing the final good, albeit at different productivities. The first is

$$h(k) = \frac{k}{q}, \quad (1)$$

which converts the capital good into the final good; capital fully depreciates when used in this technology. The second is

$$f(x, k) = zx^\alpha k^{1-\alpha}, \quad \alpha \in (0, 1), \quad (2)$$

where z is the standard neutral productivity shock and α is the share of the borrowed input. This is the technology used when lending occurs; it requires both the borrower's capital k and the amount x borrowed out of the lender's endowment of X . This production technology is risky, i.e. it is subject to an idiosyncratic shock such that with probability p the production is successful, but with the complement probability it generates zero output (therefore wasting the borrowed amount x). A fraction $\delta \in (0, 1)$ of capital depreciates when used in this technology, so that $(1 - \delta)k$ is left over.

Both agents can therefore stay in autarky, using $g(X)$ and $h(k)$ respectively, or engage in lending and borrowing, in which case the borrower produces with $f(x, k)$. Following Quadrini (2011), we assume multiple sectors: the final good can be produced with $g(X)$, $h(k)$ and $f(x, k)$, although at differing productivities. Notice that we allow all production technologies to be used in a cycle: the lender might have remaining x used in production technology g , the borrower can use the undepreciated capital from production technology f in production technology h .

The assumption that borrowing is used for working capital needs rather than for investment in physical capital is supported by a number of empirical studies. Kashyap et al. (1994), using micro US data, find that bank-leveraged firms cut their inventories significantly relative to non-bank-dependent firms in response to adverse macroeconomic conditions. Paravisini et al. (2015), using Peruvian data, find that bank-dependent firms reduced their exports by more than non-bank-dependent firms during the recent crisis, and that this occurred mostly by raising firms' variable production costs rather than by affecting their stock of fixed capital. Although firms use credit for many reasons, the primary use of short-term debt is to facilitate working capital needs rather than to replace existing physical capital or to finance expansion (Cressy and Olofsson, 1997).

3.2 The Uncertainty Structure and the Contractual Friction

At the start of the period, the realisation of the idiosyncratic shock, i.e. whether the production technology f will produce positive output, is set, but it is not observable before production takes place. However, the lender owns an audit technology which reveals it and makes it common knowledge before the lending decision is made. The audit technology comes at a fixed cost $A > 0$.

If the lender audits and the project is found to be successful, he lends without imposing collateral requirements, since the borrower is able to repay out of the output of the successful project. If the audit reveals that the project is unsuccessful, no loan is extended: the lender is indifferent between lending against collateral and not lending, while the borrower is strictly better off in autarky, since she would be certain to lose the capital pledged. Both agents then use $g(X)$ and $h(k)$ respectively.

If the lender does not audit, he might still lend, but we assume he fully secures the repayment of x by using the undepreciated borrower's capital as collateral. An unaudited borrower can therefore borrow at most

$$x \leq \frac{(1 - \delta)k}{q}, \quad (3)$$

the borrower's undepreciated capital transformed into consumption good.

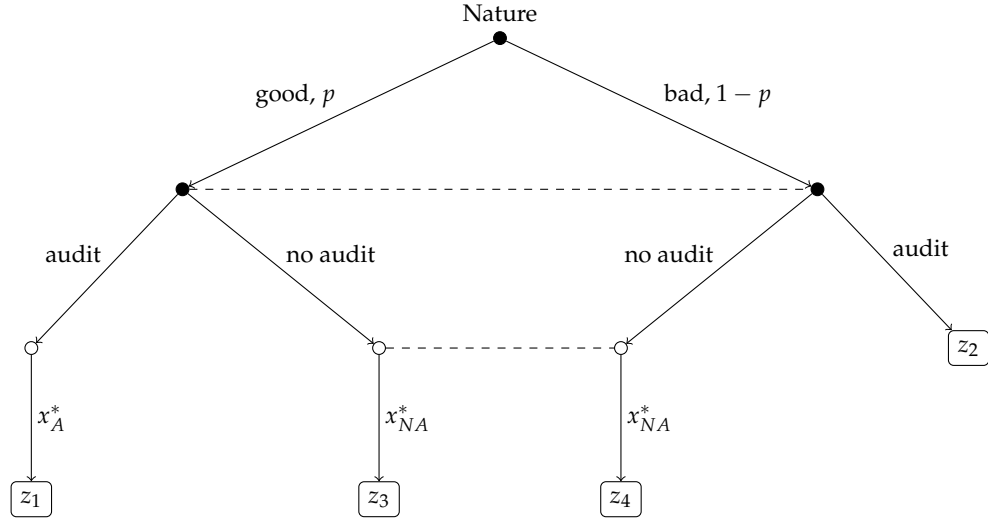
This implies that a borrower with a small capital stock may be unable to obtain her optimal borrowing demand, especially when q is high. This is in line with the evidence discussed in Section 2.1: since banks might not want to incur the audit costs when dealing with small firms, uncertainty remains about whether those firms will meet their debt obligations, and banks require high collateral as a precaution.

3.3 The Lender–Borrower Relationship

After the shocks are realised and the audit decision has been made, the matched lender and borrower may engage in a bargaining process in which the lender agrees to lend a certain amount of their endowment to the borrower, who will use these resources to engage in production using technology f . The division of output is agreed through a Nash Bargaining solution, where $\gamma \in [0, 1]$ is the lender's bargaining weight and $1 - \gamma$ the borrower's. Nash Bargaining is widely used in the macro-labour literature and also in the macroeconomic financial frictions literature in setting credit contracts (see Den Haan et al., 2003). It is worth stressing that bargaining determines the division of the surplus, that is the payment schedule from the borrower to the lender $w(x)$, while the borrower determines the size of the surplus through her choice of x .

3.4 Timing

We can describe the interaction between borrower and lender as a sequential game. The game tree would be described as follows: (i) the aggregate and idiosyncratic shocks are realised, with the realisation of the idiosyncratic shock unobservable; (ii) the lender decides whether to audit the borrower he is matched with, and if he audits, the realisation of the idiosyncratic shock becomes common knowledge; (iii) the borrower declares her optimal demand x^* , subject to the collateral constraint (3) if she has not been audited; (iv) the agents Nash bargain over the division of the surplus; (v) if lending is agreed, the contract is written and x is lent; (vi) production with $f(x, k)$ takes place, and the unaudited borrower learns the realisation of her idiosyncratic shock; (vii) repayment occurs, either out of the output of a successful project or, for the unsuccessful unaudited borrower, by converting the equivalent amount of her remaining capital; (viii) both agents use $g(X)$ and $h(k)$ to produce the final good out of what remains of $\bar{x}$ and k respectively; and (ix) consumption takes place.



Terminal payoffs, stated as (lender, borrower) in units of the final good:

$$\begin{aligned}
 z_1 &= \left(\bar{x} - A - x_A^* + w_A^*(x_A^*), f(x_A^*, k) - w_A^*(x_A^*) + \frac{(1-\delta)k}{q} \right), \\
 z_2 &= \left(\bar{x} - A, \frac{k}{q} \right), \\
 z_3 &= \left(\bar{x} - x_{NA}^* + w_{NA}^*(x_{NA}^*), f(x_{NA}^*, k) - w_{NA}^*(x_{NA}^*) + \frac{(1-\delta)k}{q} \right), \\
 z_4 &= \left(\bar{x}, \frac{(1-\delta)k}{q} - x_{NA}^* \right).
 \end{aligned}$$

Figure 1: The lender-borrower game in extensive form. Filled nodes are the lender's, hollow nodes the borrower's, and dashed lines join nodes in the same information set.

4 Equilibrium and the Audit Threshold

In this section, we solve for the subgame perfect equilibrium of the model working by backward induction: first we derive the Nash Bargaining solution under each of the two credit treatments, then the borrower's problem, the lender's problem, to ultimately define the audit threshold $\hat{k}$ that the lender's optimisation generates. Throughout, budget constraints are expressed in terms of the final consumption good, and parameters are assumed to satisfy the participation conditions collected in Appendix 6, under which both agents prefer the match to autarky.

4.1 Nash Bargaining

The lender and the borrower bargain over the (potentially uncertain) output of production, $f(x, k)$. Let w be the lender's payoff, so that $f(x, k) - w$ is the borrower's. The lender's outside option is x , the amount he would otherwise have used in his autarkic production technology g . The borrower's opportunity cost of agreement is $\delta k/q$ under audit: she retains $(1 - \delta)k$ whether or not the match takes place, so what she forgoes by producing with $f(x, k)$ rather than $h(k)$ is the fraction of capital that depreciates in the former. Under no audit she additionally forgoes the borrowed amount with probability $1 - p$, so her opportunity cost is $\delta k/q + (1 - p)x$. The surplus of the match is accordingly $f(x, k) - x - \delta k/q$ under audit and $pf(x, k) - x - \delta k/q$ under no audit.

The audit case. The bargaining problem is

$$\max_{w(x)} [w(x) - x]^\gamma \left[f(x, k) - w(x) - \frac{\delta k}{q} \right]^{1-\gamma}, \quad (4)$$

and the lender's share, that is the payment rule, is

$$w_A^*(x) = \gamma \left[f(x, k) - x - \frac{\delta k}{q} \right] + x, \quad (5)$$

while the borrower receives $f(x, k) - w_A^*(x)$.

The no audit case. Since the loan is repaid in either state but output is produced only in the successful one, the bargaining problem becomes

$$\max_{w(x)} [p(w(x) - x)]^\gamma \left[p(f(x, k) - w(x)) - (1 - p)x - \frac{\delta k}{q} \right]^{1-\gamma}, \quad (6)$$

with solution

$$w_{NA}^*(x) = x + \frac{\gamma}{p} \left[pf(x, k) - x - \frac{\delta k}{q} \right] = \gamma f(x, k) + \left(1 - \frac{\gamma}{p} \right) x - \frac{\gamma \delta k}{pq}. \quad (7)$$

4.2 The Borrower's Problem

The borrower chooses an optimal demand x to maximise $u(c) = c$, that is her expected profits in terms of the final good. Since we operate under sequential rationality, she takes the bargaining solution into account when choosing demand x . There are two versions of this problem, depending on whether the lender chose to audit or not, i.e. whether the idiosyncratic shock is common knowledge or not.

The audit case. Substituting the optimal payment rule from (5) and adding the final good produced out of the undepreciated capital, the borrower solves

$$\max_x \left[f(x, k) - w_A^*(x) + \frac{(1 - \delta)k}{q} \right] = \max_x \left[(1 - \gamma) \left(f(x, k) - x - \frac{\delta k}{q} \right) + \frac{k}{q} \right], \quad (8)$$

which yields the optimal borrowing demand under audit

$$x_A^* = (z\alpha)^{\frac{1}{1-\alpha}} k. \quad (9)$$

The no audit case. The borrower now maximises expected consumption subject to the collateral constraint (3):

$$\begin{aligned} \max_{0 \leq x \leq \frac{(1-\delta)k}{q}} & \left\{ p \left[f(x, k) - w_{NA}^*(x) + \frac{(1 - \delta)k}{q} \right] + (1 - p) \left[\frac{(1 - \delta)k}{q} - x \right] \right\} \\ &= \max_{0 \leq x \leq \frac{(1-\delta)k}{q}} \left\{ (1 - \gamma) [pf(x, k) - x] + \frac{[1 - \delta(1 - \gamma)]k}{q} \right\}, \end{aligned} \quad (10)$$

with solution

$$x_{NA}^* = \min \left\{ (pz\alpha)^{\frac{1}{1-\alpha}} k, \frac{(1 - \delta)k}{q} \right\}. \quad (11)$$

Both demands are linear in k , and $x_A^* > x_{NA}^*$ for all $k > 0$: auditing supports a strictly larger loan, both because the unaudited borrower discounts output by p or because she may be constrained.

4.3 Binding and Non-Binding Collateral Constraints

Since the collateral limit and the unconstrained demand are both proportional to k , whether the constraint binds does not depend on the borrower's size but only on the parameters. Com-

paring the two arguments in (11), the constraint binds if and only if

$$q > \bar{q} \equiv \frac{1 - \delta}{(pz\alpha)^{\frac{1}{1-\alpha}}}. \quad (12)$$

When capital is expensive, that is when q is high, the undepreciated capital that the unaudited borrower can pledge is worth little in terms of the final good, and she is unable to borrow her optimal demand. Two cases must therefore be carried through the lender's problem, and they give rise to two expressions for the threshold.

4.4 The Lender's Problem

The lender chooses between three courses of action: staying in autarky, lending with audit, and lending without audit. The problem is solved by comparing his indirect utility V^L under the three options. In autarky he uses his entire endowment to produce using technology $g(X)$, so that

$$V_{Aut}^L = \bar{x}. \quad (13)$$

Lending with audit. If the audit reveals a successful project, the lender supplies the borrower's demand and uses the remainder $(\bar{x} - A - x)$ in $g(X)$; if it reveals an unsuccessful one, no loan is extended and he uses $(\bar{x} - A)$ for autarky production. His expected utility is therefore

$$EU_A^L = p[\bar{x} - A - x + w_A(x)] + (1 - p)(\bar{x} - A), \quad (14)$$

and evaluating at x_A^* gives the indirect utility

$$V_A^L = \bar{x} - A + p\gamma k \left[z(z\alpha)^{\frac{\alpha}{1-\alpha}} - (z\alpha)^{\frac{1}{1-\alpha}} - \frac{\delta}{q} \right]. \quad (15)$$

Lending without audit. If the project succeeds the lender receives his share $w_{NA}^*(x)$ and uses $(\bar{x} - x)$ in $g(X)$; if it fails he recovers x out of the collateralised capital and uses $(\bar{x} - x + x)$. His expected utility is

$$EU_{NA}^L = p[\bar{x} - x + w_{NA}(x)] + (1 - p)(\bar{x} - x + x), \quad (16)$$

and two cases follow. When the constraint does not bind, $q \leq \bar{q}$, the borrower obtains her optimal demand and

$$V_{NA1}^L = \bar{x} + \gamma k \left[pz(pz\alpha)^{\frac{\alpha}{1-\alpha}} - (pz\alpha)^{\frac{1}{1-\alpha}} - \frac{\delta}{q} \right]. \quad (17)$$

When it binds, $q > \bar{q}$, the borrower obtains the maximum $x = (1 - \delta)k/q$ and

$$V_{NA2}^L = \bar{x} + \gamma k \left[\frac{pz(1 - \delta)^\alpha}{q^\alpha} - \frac{1}{q} \right]. \quad (18)$$

4.5 The Audit Threshold

Auditing is optimal for the lender when $V_A^L \geq V_{NAi}^L$, $i \in \{1, 2\}$. Using (15) against (17) and (18), and noting that $z(z\alpha)^{\frac{\alpha}{1-\alpha}} - (z\alpha)^{\frac{1}{1-\alpha}} = \frac{1-\alpha}{\alpha}(z\alpha)^{\frac{1}{1-\alpha}}$, we can derive two threshold levels for k .

The threshold in the non-binding case is

$$\hat{k}_{NA1} = \frac{A}{\gamma \left[\frac{1-\alpha}{\alpha}(z\alpha)^{\frac{1}{1-\alpha}} \left(p - p^{\frac{1}{1-\alpha}} \right) + \frac{\delta(1-p)}{q} \right]}, \quad (19)$$

and in the binding case

$$\hat{k}_{NA2} = \frac{A}{\gamma \left[\frac{p(1-\alpha)}{\alpha}(z\alpha)^{\frac{1}{1-\alpha}} + \frac{1-p\delta}{q} - \frac{pz(1-\delta)^\alpha}{q^\alpha} \right]}. \quad (20)$$

In particular, since the inequalities are linear in k , it is easy to conclude that the lender prefers to audit when $k \geq \hat{k}_{NAi}$, $i \in \{1, 2\}$. At this stage, we provide no comments on the sign of k and therefore whether such threshold actually exists. To simplify the analysis, we define D_1 and D_2 from the corresponding denominators in 19 and 20 as follows

$$D_1 = \frac{1-\alpha}{\alpha}(z\alpha)^{\frac{1}{1-\alpha}} \left(p - p^{\frac{1}{1-\alpha}} \right) + \frac{\delta(1-p)}{q}, \quad (21)$$

$$D_2 = \frac{p(1-\alpha)}{\alpha}(z\alpha)^{\frac{1}{1-\alpha}} + \frac{1-p\delta}{q} - \frac{pz(1-\delta)^\alpha}{q^\alpha}, \quad (22)$$

so that $\hat{k}_{NAi} = A/(\gamma D_i)$ and γD_i is the per unit of capital wedge between the lender's payoff under audit and under no audit.

Intuition. Auditing supports a strictly larger loan, $x_A^*(k) > x_{NA}^*(k)$ for all $k > 0$, and hence a strictly larger payment to the lender, $w_A^*(k) > w_{NA}^*(k)$. Against this the lender must set the fixed cost A , which does not vary with the borrower's size. Both loan demands are proportional to k , and the audited demand has the steeper slope, so the wedge in the lender's revenue between the two treatments widens as k rises while the cost of obtaining it does not. There is accordingly a level of capital at which the widening revenue wedge exactly covers A , and beyond which auditing pays. Below it, the return from the larger loan is too small to offset the audit cost, and the lender falls back on collateral. Borrowers who are identical in every respect other than their initial capital are therefore sorted into two credit treatments, which is the first

observation the model is meant to account for.

5 Properties of the Audit Threshold

This section establishes when $\hat{k}$ exists and how it responds to the model's parameters. Since $\hat{k}_{NAi} = A/(\gamma D_i)$, everything turns on the sign and the behaviour of D_i .

5.1 Existence

We first provide a proposition explaining when a threshold exists.

Proposition 1 (Existence of the threshold). *In the non-binding case, $q \leq \bar{q}$, a finite and strictly positive audit threshold exists for all admissible parameter values. In the binding case, $q > \bar{q}$, a finite and strictly positive threshold exists if and only if $D_2 > 0$; when $D_2 \leq 0$ the lender never audits, whatever the borrower's capital, and every borrower is lent to against collateral.*

Proof. Both terms of D_1 in (21) are strictly positive: the second because $p < 1$ and $\delta, q > 0$, and the first because $1/(1 - \alpha) > 1$ implies $p^{1/(1-\alpha)} < p$ for $p \in (0, 1)$. Hence $D_1 > 0$ and $\hat{k}_{NA1} = A/(\gamma D_1)$ is strictly positive.

For the binding case, $V_A^L - V_{NA2}^L = -A + \gamma k D_2$. If $D_2 \leq 0$ this is strictly negative for every $k > 0$, so no audit occurs at any size; if $D_2 > 0$ it is increasing in k and crosses zero once, at the positive threshold $\hat{k}_{NA2}$. $\square$

Sorting by size is not a universal feature of the environment. Focus on the last result: when capital is sufficiently expensive relative to the productivity of the borrower's technology, the collateral that the unaudited borrower can pledge is worth very little. Therefore, the loan the lender can safely extend is small regardless of the borrower's capital, and the revenue wedge never widens enough to justify the audit cost. In that region the segmentation between an SME and a corporate treatment disappears and every firm is treated in the same way.

5.2 Comparative Statics

The following proposition provides the direction of changes to the audit thresholds $\hat{k}$ with respect to the model parameters.

Proposition 2. *In the non-binding case, $\partial \hat{k}_{NA1}/\partial A > 0$, $\partial \hat{k}_{NA1}/\partial \gamma < 0$, $\partial \hat{k}_{NA1}/\partial z < 0$, $\partial \hat{k}_{NA1}/\partial \delta < 0$ and $\partial \hat{k}_{NA1}/\partial q > 0$, for any combination of parameters. In the binding case, provided $D_2 > 0$, $\partial \hat{k}_{NA2}/\partial A > 0$, $\partial \hat{k}_{NA2}/\partial \gamma < 0$, $\partial \hat{k}_{NA2}/\partial z < 0$ and $\partial \hat{k}_{NA2}/\partial \delta < 0$, where the last two follow from $q > \bar{q}$ itself. The signs of $\partial \hat{k}_{NA1}/\partial p$, $\partial \hat{k}_{NA2}/\partial p$ and $\partial \hat{k}_{NA2}/\partial q$ are not determined by the model.*

The derivations are collected in Appendix 6. Table 1 summarises. Two features deserve comment.

First, four of the six parameters act unambiguously. A higher audit cost raises the threshold and shrinks the set of firms receiving corporate treatment; a higher bargaining weight for the lender lowers it, since he appropriates more of the additional surplus that auditing creates, and γ enters only multiplicatively; a higher neutral productivity raises the return on the larger audited loan and so lowers the threshold; and a higher depreciation rate lowers it as well, since it decreases the collateral that makes the unaudited alternative attractive.

Second, the signs of z and δ in the binding case require no additional restriction. Both reduce to the requirement $(z\alpha)^{1/(1-\alpha)} > (1-\delta)/q$, which is implied by the definition of the binding region itself: $q > \bar{q}$ means $(pz\alpha)^{1/(1-\alpha)} > (1-\delta)/q$, and $(z\alpha)^{1/(1-\alpha)} > (pz\alpha)^{1/(1-\alpha)}$ since $p < 1$. Being in the binding case thus delivers the two signs.

Table 1: Comparative statics of the audit threshold

Parameter	$\hat{k}_{NA1}$	$\hat{k}_{NA2}$	Notes
A	+	+	
γ	−	−	
z	−	−	$\hat{k}_{NA2}$: follows from $q > \bar{q}$
δ	−	−	$\hat{k}_{NA2}$: follows from $q > \bar{q}$
q	+	?	
p	?	?	non-monotone; see Section 5.3

Entries report the sign of $\partial \hat{k} / \partial (\cdot)$. Signs for $\hat{k}_{NA2}$ are conditional on $D_2 > 0$, that is on the threshold existing at all (Proposition 1). A question mark indicates that the sign is not determined by the model and depends on the configuration of parameters.

5.3 The Non-Monotonicity in p

We close this section by discussing the result on the non monotonicity in p .

Consider D_1 in (21). The first term is proportional to $p - p^{1/(1-\alpha)}$, which vanishes at both $p = 0$ and $p = 1$ and is strictly positive and hump-shaped in between; the second term, $\delta(1-p)/q$, declines linearly to zero. Hence $D_1 \rightarrow \delta/q > 0$ as $p \rightarrow 0$, so that $\hat{k}_{NA1} \rightarrow Aq/(\gamma\delta)$, while $D_1 \rightarrow 0$ as $p \rightarrow 1$, so that $\hat{k}_{NA1}$ diverges. Whenever D_1 is increasing at low p , the threshold is non-monotone, falling and then rising, with an interior minimum.

Raising the success probability does two things which might not point in the same direction. For high values of p , on one hand increasing p raises the scale of the unaudited borrower's loan relative to the audited one, since $x_{NA}^*/x_A^* = p^{1/(1-\alpha)}$ rises towards unity, and it simultaneously shrinks the expected loss the lender avoids by not lending to a project the audit would have revealed as bad. Both effects reduce the advantage of auditing and therefore push $\hat{k}$ up. On the other hand, when p is low, higher chances of successful production raise the return on the larger audited loan by more than they raise the return on the unaudited one, widening the

advantage and pulling $\hat{k}$ down. The first force dominates near $p = 1$, where there is little left for an audit to discover and the threshold diverges, i.e. non auditing is the best option for most firms; the second dominates at low p , where learning the realisation of the production shock gets more valuable. Figure 2 illustrates this pattern.

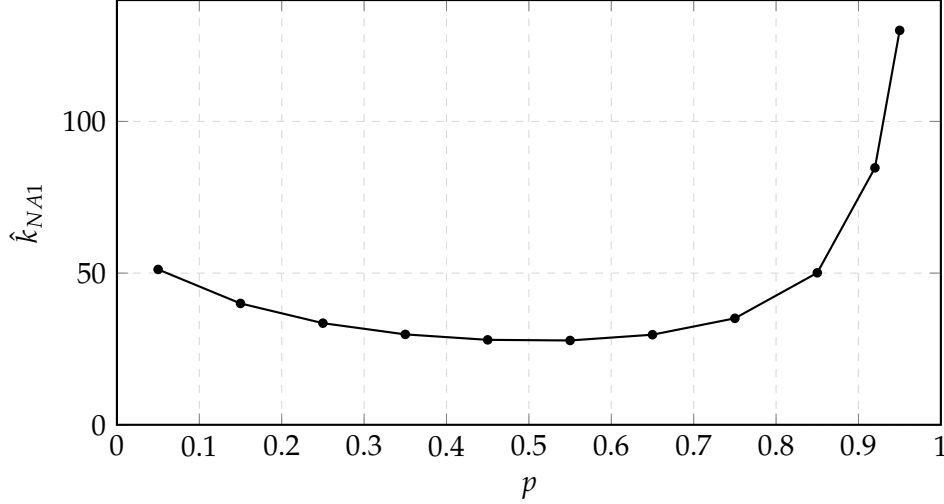


Figure 2: The audit threshold against the success probability, non-binding case (guaranteed by parameters). Parameters: $\alpha = 2/3$, $\delta = 0.1$, $z = 1$, $q = 3$, $\gamma = 0.5$, $A = 1$. The threshold falls and then rises, with a minimum near $p \approx 0.5$, and diverges as $p \rightarrow 1$.

6 Discussion and Next Steps

This draft includes a first set of results in our research agenda.

The model as it stands accounts for the first of the two observations set out in Section 1. Firms which differ only in the capital they own are sorted endogenously into two credit treatments by the lender's own optimisation: those above $\hat{k}$ are evaluated and lend without collateral, those below are not evaluated and lend against it. No exogenous difference in type, technology or project quality is required to produce the segmentation that Albareto et al. (2011) and Binks and Ennew (1997) document.

The second observation, the non-linear decline in investment by size reported in Crawford et al. (2013) and Audretsch and Elston (2002), requires more than the threshold itself. The ingredient the present setting lacks is an object of investment: the loan x finances working capital, and the borrower's capital stock is given within the period, so the model has a great deal to say about who receives which contract and nothing yet to say about how much any of them invests.

The natural extension to this framework is the one to two periods, with the lender endowed in each period but the borrower not, which would force the borrower to invest in order to con-

sume in the second period and would provide an alternative route to the investment margin. The comparative statics above are then at the core of the argument: an adverse shock to z raises $\hat{k}$, so that borrowers in a band immediately below the new threshold lose the audited treatment they previously enjoyed and suffer a discrete deterioration in credit terms on top of the direct effect of the shock, while borrowers well above and well below the threshold experience only the direct effect. The shape of the resulting cross-section in size is what the data are said to show. Since z and q move $\hat{k}$ through different channels, their relative magnitudes will then matter for the shape.

Several other extensions might be worth pursuing. Removing the collateral constraint altogether should still deliver a threshold, though a higher one, since the audited and unaudited loans would then differ only through p ; this would isolate how much of the sorting is driven by the collateral channel as opposed to the information channel. Allowing the audit cost to have a variable component increasing in k , rather than being purely fixed, is closer to how evaluation costs are actually incurred and would flatten the revenue wedge that generates the threshold, possibly bounding the set of audited firms from above as well as below. The model is also grounded on assumption of risk neutrality of agents; introducing risk aversion would change the bargaining solution and is likely to interact with the non-monotonicity in p discussed in Section 5.3. The informational structure at present is symmetric, and a version in which the borrower knows more than the lender would connect the mechanism more directly to the screening literature discussed in Section 2.1. The longer term aim is to build a component that can be plugged into a larger macroeconomic model, an overlapping generations structure being the natural candidate, in which the distribution of firms around an endogenous and shifting audit threshold is the mechanism through which aggregate shocks are transmitted unevenly across the size distribution.

References

- Albareto, G., M. Benvenuti, S. Mocetti, M. Pagnini, and P. Rossi (2011). "The organization of lending and the use of credit scoring techniques in Italian banks." *Journal of Financial Transformation*, 32, 143–168.
- Audretsch, D. B., and J. A. Elston (2002). "Does firm size matter? Evidence on the impact of liquidity constraints on firm investment behavior in Germany." *International Journal of Industrial Organization*, 20(1), 1–17.
- Ayyagari, M., A. Demirgüç-Kunt, and V. Maksimovic (2011). "Small vs. young firms across the world: contribution to employment, job creation, and growth." World Bank Policy Research Working Paper 5631.

- Beck, T. (2013). "Bank financing for SMEs: lessons from the literature." *National Institute Economic Review*, 225(1), R23–R38.
- Becsi, Z., V. E. Li, and P. Wang (2005). "Heterogeneous borrowers, liquidity, and the search for credit." *Journal of Economic Dynamics and Control*, 29(8), 1331–1360.
- Benito, A., K. Neiss, S. Price, and L. Rachel (2010). "The impact of the financial crisis on supply." *Bank of England Quarterly Bulletin*, Q2.
- Bernanke, B., and M. Gertler (1989). "Agency costs, net worth, and business fluctuations." *American Economic Review*, 79(1), 14–31.
- Bernanke, B., M. Gertler, and S. Gilchrist (1996). "The financial accelerator and the flight to quality." *Review of Economics and Statistics*, 78(1), 1–15.
- Bernanke, B., M. Gertler, and S. Gilchrist (1999). "The financial accelerator in a quantitative business cycle framework." In *Handbook of Macroeconomics*, Vol. 1C, 1341–1393.
- Besanko, D., and A. V. Thakor (1987). "Collateral and rationing: sorting equilibria in monopolistic and competitive credit markets." *International Economic Review*, 28(3), 671–689.
- Bester, H. (1985). "Screening vs. rationing in credit markets with imperfect information." *American Economic Review*, 75(4), 850–855.
- Binks, M. R., and C. T. Ennew (1997). "Small business and relationship banking: the impact of participative behavior." *Entrepreneurship Theory and Practice*, 21(4), 83–92.
- BVA BDRC (2026). *SME Finance Monitor: Q4 2025 Report*. March 2026.
- Caminal, R., and C. Matutes (1997). "Bank solvency, market structure, and monitoring incentives." CEPR Discussion Paper 1665.
- Carlstrom, C. T., and T. S. Fuerst (1997). "Agency costs, net worth, and business fluctuations: a computable general equilibrium analysis." *American Economic Review*, 87(5), 893–910.
- Chodorow-Reich, G. (2014). "The employment effects of credit market disruptions: firm-level evidence from the 2008–9 financial crisis." *Quarterly Journal of Economics*, 129(1), 1–59.
- Competition and Markets Authority (2016). *Retail Banking Market Investigation: Final Report*.
- Cowan, C. D., and A. M. Cowan (2006). *A Survey Based Assessment of Financial Institution Use of Credit Scoring for Small Business Lending*. US Small Business Administration.
- Crawford, C., W. Jin, and H. Simpson (2013). "Firms' productivity, investment and training: what happened during the recession and how was it affected by the National Minimum Wage?" IFS Report R76.

- Cressy, R., and C. Olofsson (1997). "European SME financing: an overview." *Small Business Economics*, 9(2), 87–96.
- Department for Business and Trade (2025). *Business Population Estimates for the UK and Regions 2025*. UK Government.
- Den Haan, W. J., G. Ramey, and J. Watson (2003). "Liquidity flows and fragility of business enterprises." *Journal of Monetary Economics*, 50(6), 1215–1241.
- Duygan-Bump, B., A. Levkov, and J. Montoriol-Garriga (2015). "Financing constraints and unemployment: evidence from the Great Recession." *Journal of Monetary Economics*, 75, 89–105.
- Fisher, J. D. M. (1997). "Relative prices, complementarities and comovement among components of aggregate expenditures." *Journal of Monetary Economics*, 39(3), 449–474.
- Fisher, J. D. M. (2006). "The dynamic effects of neutral and investment-specific technology shocks." *Journal of Political Economy*, 114(3), 413–451.
- Fort, T. C., J. Haltiwanger, R. S. Jarmin, and J. Miranda (2013). "How firms respond to business cycles: the role of firm age and firm size." *IMF Economic Review*, 61(3), 520–559.
- Greenwood, J., Z. Hercowitz, and G. W. Huffman (1988). "Investment, capacity utilization, and the real business cycle." *American Economic Review*, 78(3), 402–417.
- Greenwood, J., Z. Hercowitz, and P. Krusell (1997). "Long-run implications of investment-specific technological change." *American Economic Review*, 87(3), 342–362.
- Greenwood, J., Z. Hercowitz, and P. Krusell (2000). "The role of investment-specific technological change in the business cycle." *European Economic Review*, 44(1), 91–115.
- Haltiwanger, J., R. S. Jarmin, and J. Miranda (2010). "Who creates jobs? Small vs. large vs. young." NBER Working Paper 16300.
- Hart, O., and J. Moore (1994). "A theory of debt based on the inalienability of human capital." *Quarterly Journal of Economics*, 109(4), 841–879.
- Jaffee, D. M., and T. Russell (1976). "Imperfect information, uncertainty, and credit rationing." *Quarterly Journal of Economics*, 90(4), 651–666.
- Justiniano, A., G. E. Primiceri, and A. Tambalotti (2010). "Investment shocks and business cycles." *Journal of Monetary Economics*, 57(2), 132–145.
- Justiniano, A., G. E. Primiceri, and A. Tambalotti (2011). "Investment shocks and the relative price of investment." *Review of Economic Dynamics*, 14(1), 102–121.

- Kashyap, A. K., O. A. Lamont, and J. C. Stein (1994). "Credit conditions and the cyclical behavior of inventories." *Quarterly Journal of Economics*, 109(3), 565–592.
- Kiyotaki, N., and J. Moore (1997). "Credit cycles." *Journal of Political Economy*, 105(2), 211–248.
- Neumark, D., B. Wall, and J. Zhang (2011). "Do small businesses create more jobs? New evidence for the United States from the National Establishment Time Series." *Review of Economics and Statistics*, 93(1), 16–29.
- Paravisini, D., V. Rappoport, P. Schnabl, and D. Wolfenzon (2015). "Dissecting the effect of credit supply on trade: evidence from matched credit-export data." *Review of Economic Studies*, 82(1), 333–359.
- Petersen, M. A., and R. G. Rajan (1994). "The benefits of lending relationships: evidence from small business data." *Journal of Finance*, 49(1), 3–37.
- Quadrini, V. (2011). "Financial frictions in macroeconomic fluctuations." *Federal Reserve Bank of Richmond Economic Quarterly*, 97(3), 209–254.
- Şahin, A., S. Kitao, A. Cororaton, and S. Laiu (2011). "Why small businesses were hit harder by the recent recession." *Current Issues in Economics and Finance*, 17(4), Federal Reserve Bank of New York.
- Smets, F., and R. Wouters (2007). "Shocks and frictions in US business cycles: a Bayesian DSGE approach." *American Economic Review*, 97(3), 586–606.
- Villas-Boas, J. M., and U. Schmidt-Mohr (1999). "Oligopoly with asymmetric information: differentiation in credit markets." *RAND Journal of Economics*, 30(3), 375–396.

Appendix

Participation Conditions For lending and borrowing to take place rather than both agents remaining in autarky, the technology $f(x, k)$ must be sufficiently productive relative to $g(X)$ and $h(k)$. Because the Nash Bargaining solution gives each agent a fixed share of the match surplus, the lender's and the borrower's participation constraints coincide: both are satisfied exactly when the surplus of the match is strictly positive. This reduces the conditions to one per case.

Under audit, the borrower's payoff net of her outside option is $(1 - \gamma) [f(x, k) - x - \delta k/q]$ and the lender's is $\gamma [f(x, k) - x - \delta k/q]$. Evaluating at x_A^* and using $f - x = \frac{1-\alpha}{\alpha} (z\alpha)^{1/(1-\alpha)} k$, both are strictly positive if and only if

$$\frac{1 - \alpha}{\alpha} (z\alpha)^{\frac{1}{1-\alpha}} > \frac{\delta}{q}. \quad (23)$$

Under no audit with a non-binding constraint, the corresponding surplus is $pf(x, k) - x - \delta k/q$, evaluated at $x_{NA}^* = (pz\alpha)^{1/(1-\alpha)}k$, so that

$$\frac{1-\alpha}{\alpha}(pz\alpha)^{\frac{1}{1-\alpha}} > \frac{\delta}{q}. \quad (24)$$

Note that (24) implies (23), since $p < 1$. Under no audit with a binding constraint, substituting $x = (1-\delta)k/q$ gives

$$pz(1-\delta)^\alpha q^{1-\alpha} > 1. \quad (25)$$

Throughout the paper we maintain (24) when $q \leq \bar{q}$ and (23) together with (25) when $q > \bar{q}$. No separate condition is needed to ensure that the lender prefers auditing to autarky for $k \geq \hat{k}$: in that region $V_A^L \geq V_{NAi}^L$ by construction, and $V_{NAi}^L \geq V_{Aut}^L$ by the conditions above.

Derivations for Proposition 2 Throughout, $\hat{k}_{NAi} = A/(\gamma D_i)$ with D_i as defined in (21) and (22), and $D_i > 0$ in the relevant region by Proposition 1. Since $\partial \hat{k}/\partial \theta = -(A/\gamma D_i^2) \partial D_i/\partial \theta$ for any parameter θ other than A and γ , the sign of $\partial \hat{k}/\partial \theta$ is the opposite of the sign of $\partial D_i/\partial \theta$.

A and γ . Both enter only outside D_i , giving $\partial \hat{k}/\partial A = 1/(\gamma D_i) > 0$ and $\partial \hat{k}/\partial \gamma = -A/(\gamma^2 D_i) < 0$.

Non-binding case. Write $B \equiv \frac{1-\alpha}{\alpha}(z\alpha)^{1/(1-\alpha)} > 0$. Then $\partial D_1/\partial z > 0$, since B is increasing in z and the second term of D_1 does not involve z ; $\partial D_1/\partial \delta = (1-p)/q > 0$; and $\partial D_1/\partial q = -\delta(1-p)/q^2 < 0$. The corresponding signs for $\hat{k}_{NA1}$ are therefore negative, negative and positive. For p ,

$$\frac{\partial D_1}{\partial p} = B \left(1 - \frac{1}{1-\alpha} p^{\frac{\alpha}{1-\alpha}} \right) - \frac{\delta}{q},$$

whose sign is not determined: the bracket is positive for small p and negative for p close to one, since $\frac{1}{1-\alpha} p^{\alpha/(1-\alpha)} \rightarrow \frac{1}{1-\alpha} > 1$ as $p \rightarrow 1$, while $-\delta/q < 0$ throughout.

Binding case. Differentiating (22),

$$\frac{\partial D_2}{\partial z} = p \left[(z\alpha)^{\frac{\alpha}{1-\alpha}} - \left(\frac{1-\delta}{q} \right)^\alpha \right], \quad \frac{\partial D_2}{\partial \delta} = p \left[\frac{z\alpha(1-\delta)^{\alpha-1}}{q^\alpha} - \frac{1}{q} \right].$$

Both brackets are positive if and only if $(z\alpha)^{1/(1-\alpha)} > (1-\delta)/q$. This holds throughout the binding region: $q > \bar{q}$ is equivalent to $(pz\alpha)^{1/(1-\alpha)} > (1-\delta)/q$, and $(z\alpha)^{1/(1-\alpha)}$ exceeds $(pz\alpha)^{1/(1-\alpha)}$ because $p < 1$. Hence $\partial \hat{k}_{NA2}/\partial z < 0$ and $\partial \hat{k}_{NA2}/\partial \delta < 0$ unconditionally within that region. For the remaining two parameters,

$$\frac{\partial D_2}{\partial p} = \frac{1-\alpha}{\alpha}(z\alpha)^{\frac{1}{1-\alpha}} - \frac{\delta}{q} - \frac{z(1-\delta)^\alpha}{q^\alpha}, \quad \frac{\partial D_2}{\partial q} = -\frac{1-p\delta}{q^2} + \frac{\alpha pz(1-\delta)^\alpha}{q^{\alpha+1}},$$

neither of which has a determinate sign.