

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1100

Note: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than mere acknowledgment by a writer that he has access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

BANKS VERSUS BONDS:
A SIMPLE THEORY OF COMPARATIVE FINANCIAL INSTITUTIONS

Sandeep Baliga and Ben Polak

May 1995

BANKS VERSUS BONDS:

A Simple Theory of Comparative Financial Institutions.*

By

Sandeep Baliga[†] and Ben Polak[‡]

This Draft: 27 April 1995

(First Draft: 29 November 1993)

* We thank Stephen King and Andreu Mas-Colell, and also Abhijit Banerjee, Herbert Bloch, Margaret Bray, Jorg Decressin, Jeremy Edwards, John Fernald, Simon Grant, Tim Guinnane, Al Klevorick, Eric Maskin, Boaz Moselle, Emma Rothschild, Paola Sapienza, Kotaro Suzumura, Tomas Sjöström, David Weinstein, Jeff Williamson, Ernest von Thadden, Yishay Yafeh, Oved Yosha, and Richard Zeckhauser and seminar participants in Cambridge, ECARE, Stanford, and Wharton for helpful comments and discussions.

[†] Kings College, Cambridge

[‡] Yale University.

Abstract.

We use a simple, graphical moral hazard model to compare monitored bank lending versus non-monitored bond issues as sources of external funds for industry. We contrast the conditions that theoretically favour each system, such as the size and number of firms, with conditions prevailing when these financial systems were developed during the British and German Industrial Revolutions. Then, to address the question why different systems have persisted, we embed the model in an entry game in which firm size and number are endogenous. We show that multiple equilibria can exist if financiers take the industrial structure as given and vice versa. Finally, we compare these equilibria in welfare terms.

(JEL # N20, D82, G20).

1. INTRODUCTION AND MOTIVATION.

Three questions motivate this paper. First, why did different methods of finance emerge from the British and German industrial revolutions? Banks figured prominently in Germany; notably, but not only, the *Grossbanken*. External finance, to the degree it was used at all in England, was often in the form of tradeable bills of exchange or promissory notes.¹ Second, why did these two modes of finance not converge more quickly over time? Separate so-called German and Anglo-Saxon financial systems persist, albeit with some changes, today. And third, does it matter?

Each of the three questions above has a good pedigree. For example, it was the power and importance of the German universal banks that led Hilferding to develop his theory of *Finance Capital*.² Later, different methods of finance were among the main contrasts identified by Gerschenkron in his seminal comparative study of early and late industrial revolutions:

"The industrialisation of England had proceeded without any substantial utilisation of banking for ... investment purposes. ... [Whereas] the continental practices in the field of industrial investment banking must be conceived as specific instruments of industrialisation in a backward country."³

More recently, in his survey of new thinking on the British industrial revolution, Mokyr wrote of industrial banks: "why such institutions were relatively unimportant is still an unanswered problem."⁴

The third question--the pros and cons of the Anglo-Saxon versus the German financial system--has given rise to much heated debate. The grass has often seemed greener on the other side of the fence. Analyses of Anglo-Saxon economies often blame the supposed greater separation of industry from finance for the decline of Britain from the late 19th century onward and of the US today.⁵ Analyses of Germany sometimes argue that *Grossbanken* hindered growth, while similar complaints are sometimes made of *Zaibatsu* and *Keiretsu* in

¹ See, for example: Tilly (1992), Crouzet (1963) and Neal (1994). Edwards and Ogilvie (1995) warn against exaggerating the role of universal banks in Germany.

² Hilferding, R. (1910).

³ Gerschenkron, A, (1962, p. 14)

⁴ Mokyr, J., (1985, p. 37).

⁵ For example, Best and Humphries (1986, p. 223) write of Britain: "The lack of integration between finance and industry adversely affected the volume and allocation of British industrial investment and the long-term competitive performance of British industry compared with its international rivals." While Calomiris (1992, pp. 1-2) writes of the US: "Large-scale industrial investment was stunted relative to its potential by a faulty financial system." See also, for example, Ingham (1984); and Kennedy (1987 and 1990). For a critiques of the US and UK financial systems today see Mayer (1991) and Porter (1992).

Japan.⁶ Most recently, the debate has resurfaced in the guise of choosing appropriate financial institutions for Eastern Europe.

One major difference between the two financial systems was the degree to which creditors monitored firms. German industrial banks "established the closest possible relations with industrial enterprises."⁷ J. Riesser, the self-appointed spokesman for industrial bankers, wrote:

"Sooner or later this connection finds further expression in the appointment of members of the banks directorate to the supervisory council of the industrial enterprise.... it [is] the duty of the bank according to well established and sound practice of German banking, to retain such permanent control."⁸

By 1903, the six largest Berlin banks controlled 751 positions on Boards of directors.⁹ In England, creditors often preferred a "hands-off" approach. Final holders of widely traded securities may have been too distant to monitor directly. It may also have been difficult to recover the costs of monitoring when an asset was transferred. But, even when loans were direct from banks, their monitoring did not impress Riesser:

"the [English] banks have never shown any interest in the newly founded companies or in the securities issued by these companies, while it is a distinct advantage of the German system, that the German banks, even if only in the interests of their own issue credit, have been keeping a continuous watch over the development of the companies, which they founded."¹⁰

In this paper, we model both the choice and welfare consequences of different methods of financing firms, focusing on monitoring and its effect on moral hazard. This is not to imply that other issues are unimportant when comparing German and Anglo-Saxon finance systems. For example, Calomiris (1992) and Calomiris and Raff (1993) stress economies of scope enjoyed by universal banks. Dewatripont and Maskin (1990) emphasize a trade off between the ability to finance large projects and the ability credibly to commit not to refinance. Cantillo (1994) following Diamond (1984) stresses the ability of intermediaries to verify bankruptcy. Allen and Gale (1995(a), (b) and (c)) compare the ability of the two systems to smooth income intertemporally and diversify individual and aggregate risks. Hölmström (1993) uses a model similar to that presented in the first

⁶ See, for example, Neuburgher and Stokes (1974); Weinstein and Yafeh [1994]; and Masuyama (1991).

⁷ Gerschenkron (1962) p. 14.

⁸ Riesser (1909).

⁹ Hilferding, (1910) p 398. Riesser (1909) reports that 8 banks held 819 such posts in 1905. See his extensive discussion, pp. 897-920. Fohlin (1993), however, argues that representation by banks on supervisory boards was much lower before 1900. Edwards and Fischer(1994) and Edwards and Ogilvie (1995) argue that the influence of such boards both today and historically has been exaggerated.

¹⁰ Riesser (1909) p. 555.

part of section 2 to compare the benefits of both types of finance over the life-cycle of the firm. Scherfke (1993) uses a dynamic model to address similar issues. Aoki (1993) considers repeated moral hazard in teams and shows that banks can help by eliminating budget balance.

We believe that each of these issues plays a large role in the complete story. We abstract from them here only to keep our models simple. In each model, entrepreneurs borrow capital and either borrowers or lenders choose whether the entrepreneurs' actions should be directly monitored. Without monitoring, optimal contracts will induce only second best actions. For simplicity, we assume that monitoring always induces the first best actions and that there are never any problems in monitoring the monitor. Thus, participants compare the deadweight loss at the second best with the cost of monitoring.

In section 2, as a benchmark, we first present a very simple moral hazard model. A novel feature of the model is that we are able to illustrate most of its results graphically. We then compare the comparative statics of the model with the corresponding features of the British and German 18th and 19th century economies. In particular, we show that, *ceteris paribus*, increasing base interest rates favours German-style monitored loans. Shifting the bargaining power from borrowers to lenders also favours monitoring. On the first point, German interest rates were always higher than English rates, but German industrialisation took place at a time of historically low interest rates. Thus, while low rates may partially explain the failure of Britain to switch to a German financial system, they do not explain the adoption of these systems in the first place. On the second point, there is some consensus among historians that financiers were relatively more powerful vis a vis industry in Germany than in England, at least in the early stages of their respective industrial revolutions. Thus, this may form part of an explanation why a different system emerged in Germany.

We then add a little structure by explicitly modelling the costs of monitoring and the cost of tradeable debt (which we assume can not be monitored). In particular, we assume that there are internal economies of scale in monitoring with respect to the size of the loan, and external economies of scale in the costs of tradeable debt with respect to the size of the secondary asset market. Both of these assumptions seem reasonable. Monitoring one large plant is likely to be less costly than monitoring two small plants. Tradeable securities are likely to be more liquid if markets are thicker. Such external economies can arise if there are fixed costs in running a market or if matching problems contribute to transaction costs.¹¹ We also make the unrealistic but simplifying assumption

¹¹ See Diamond (1982). Strictly, these are not separate reasons. It is only worthwhile having specialised traders for a particular kind of asset if the market is thick. One role of these traders is to alleviate matching problems.

that each firm is either funded with monitored debt, or with tradeable debt, or with debt that is neither monitored nor traded, but that firms cannot have mixed capital structures.¹²

Under these assumptions, loosely speaking, firms need to be large to justify monitoring and the secondary market needs to be large to justify tradeable debt. Thus, an economy with a small number of large firms will favour monitored loans. An economy with a small number of small firms will favour non-monitored non-tradeable loans. And an economy with a large number of medium sized firms will favour tradeable non-monitored loans. This fits the stylised facts. The English industrial revolution of the late 18th century was led by industries such as textiles in which there were very many, very small firms. By the time Germany industrialised in the late 19th century, however, not only had scale increased in almost all industries but also the new leading sectors were now large scale and relatively concentrated heavy industries such as steel and chemicals. Moreover, within Germany, industrial bank credit was concentrated on large firms in heavy industries.

Two other useful results emerge from this simple model. First, given the external economies of the market, tradeable debt becomes cheaper for industry as more non-industrial securities are traded. Britain already had well developed secondary markets in government debt instruments and merchant paper by the end of the 18th century. Second, very loosely speaking, entrepreneurs having greater private wealth can also favour non-monitored loans although it can also hinder market development. Many early British entrepreneurs had some capital of their own, often built up from artisan production or trade within the textile sector.

Coordination failures can occur even in this simple model. When there are many firms the Pareto preferred equilibrium is for all firms to issue tradeable debt, but there is also an equilibrium in which all firms use non-tradeable loans. No firm will deviate since it would have to carry the full costs of the market at small scale. This may justify government policy to help promote the market either by the implicit subsidy of issuing tradeable government debt or by inhibiting universal banking by legal restrictions or taxes.

Section 3 addresses the long-run persistence of different financial systems in neighbouring economies. While the assumption that the size and the number of firms is exogenous may be a reasonable approximation when considering first forming a financial system in the midst of an industrial revolution, it is less useful when we move to consider why such systems have persisted over centuries. Therefore, we present a model in

¹² See Hölmström (1993) for a fuller treatment of capital structure.

which both the financial system and the size and number of firms is endogenous.¹³ To construct this, we embed our simple model in a larger competitive entry game. We examine two types of equilibrium: one type with entrepreneurs establishing small firms and lenders setting up institutions to handle tradeable debt; the other type with large firms and investment banks that monitor loans. We think of these as the Anglo-Saxon and German types of equilibrium.

Both Anglo-Saxon and German equilibria can exist under some parameters; that is there are multiple equilibria. The intuition for this is that industrialists may take the financial system as given when they organise firms and lenders may take the structure of industry as given when they organise financial institutions. As in the previous model, agents would have to coordinate to establish a market for tradeable debt. Thus, a German system can be sustained even when an Anglo-Saxon system, if established, would deliver goods more cheaply. In addition, however, the economy can now get stuck in an Anglo-Saxon equilibrium, even when a German system would deliver goods more cheaply. There may be no incentive for a German-style bank to enter an Anglo-Saxon economy with many small firms since large scale is required to warrant monitoring. But, since there are no German banks, there is no incentive to establish large firms. Although we emphasise scale, the general idea of such a coordination failure could be extended to other features of firms' organisation such as their accounting system that might favour either German or Anglo-Saxon forms of finance.

There are two main lessons from Section 3. First, the existence of both types of equilibrium in a model with competitive entry may explain the persistence of both financial systems. Although this is a static model, we might think of it as at least a metaphor for path dependence. Second, we show that, under some parameters at which both types of equilibrium exist, the Anglo-Saxon Pareto dominates the German equilibrium. But under other such parameters this is not the case. In particular, consumers might prefer a German equilibrium while firms prefer the Anglo-Saxon. Nothing ensures that an economy will end up in the equilibrium that is best given its characteristics. The commentators on either side of the financial fence who believe that the grass is greener on the other side, might both be right. Alternatively, of course, they might both be wrong. More generally, since not all welfare comparisons are by Pareto dominance (that is, there are both winners and losers), it is not surprising to find people in both systems who would prefer the other.

This paper is mostly about the past, but let us move briefly to the future. The

¹³ It would be nice if the wealth of entrepreneurs was also endogenous over time. Our one-shot model cannot capture this effect, but see Hölmström (1993) and Scherfke (1993).

simple models presented here suggest a "horses for courses" approach to the creation of capital markets in the formerly centrally planned economies. Where market size is potentially large and firms small, it may be optimal to create Anglo-Saxon style capital markets; where there will be fewer large-scale firms, a system with monitored debt and German style investment banks may be better. A blanket policy of imposing one system for all economies may be harmful. In deciding which system suits where, it is probably wrong to treat an economy as a *tabula rasa*. The inherited industrial structure of each particular economy--its history if you like--may still matter.

One policy that may alleviate such problems is the opening of capital markets between countries. Firms would then have access to a wide range of ways to raise capital and this may make it harder to get stuck in a bad equilibrium.¹⁴ It will be interesting to see which financial system wins out if the European Community is successful in creating a unified market in financial services.

2. A SIMPLE MORAL HAZARD MODEL

We begin by presenting a simple model that illustrates some of the trade-offs involved in choosing either the monitored bank loans often seen in the German industrial revolution, or the non-monitored (possibly tradeable) debt preferred in England. We use a standard simplified moral hazard model:¹⁵ in particular, we abstract from differences in the quality and length of loan projects in order to emphasize monitoring and issues arising from hidden actions. Where possible, we present results graphically.

2.1 The Basic Model and Its Comparative Statics.

2.1.1 The Basic Model Consider the choices facing entrepreneurs each of whom has access to a new project that requires external financing in period 1. If a project of size q is successful, it yields revenues of Pq in period 2. Assume that the price $P > 1$. If it fails, it yields zero. In either case, the project terminates in period 2. The probability of a project being successful depends only on actions taken by the entrepreneur. That is, we assume that the projects are of identical quality *ex ante*. Some of the actions or "efforts" that raise the probability of success involve private non-pecuniary costs for the entrepreneur. For example, she may have to sacrifice status or incur wrath by disciplining her employees. Let π denote the probability of success

¹⁴ But this is very loose. First, the larger integrated economy may get stuck at a bad equilibrium. Second, what is in the interest of borrowing firms may not always be in the interest of consumers.

¹⁵ The set-up is a simplified version of a part of Dewatripont and Maskin's (1990) model.

and, for a project of size q , let $q\varphi(\pi)$ be the associated private cost to the entrepreneur where $\varphi(0) = \varphi'(0) = \varphi''(0) = 0$, $\varphi(1) = \varphi'(1) = \varphi''(1) = \infty$, and $\varphi''' > 0$.

Assume that a project of size q requires the entrepreneur's participation and q units of capital to be undertaken and assume that at least part of this capital has to be borrowed. That is, if w is the entrepreneur's initial wealth, then $q > w$. Borrowing can be either in the form of a monitored or a non-monitored loan. For simplicity, we assume that partial monitoring is not allowed. If loans are not monitored, then the privately costly actions of the entrepreneur that affect the probabilities of success, π , are unobservable to lenders and will depend on incentives provided by the optimal loan contract. These incentives will, in general, only induce second-best success probability levels. If loans are monitored then such actions and hence the success probabilities they generate are observable and verifiable. That is, the advantage of monitoring is that it allows probabilities of success levels to be contracted upon and thus first-best levels to be enforced. The disadvantage of monitoring is that it is costly. We assume that there are no moral hazard problems in monitoring the monitor: perhaps, for example, the monitor has her own funds invested.

We use the subscripts A and G to denote the Anglo-Saxon and German financial systems respectively. Accordingly, let the marginal cost to the lender(s) of providing one unit of capital to the entrepreneur be given by $c \in \{c_A, c_G\}$ where c_A is the cost without monitoring and c_G is the cost with monitoring and $c_A, c_G > 1$. We will give more structure to these costs below but, for now, we assume that there are constant marginal costs. For a project of size q , let $q(1+r)$ be the the amount repaid by the entrepreneur to the lender(s) if the project is successful and qd be the amount paid if it fails. We can think of d as a per-unit default payment, and of r , in the special case where the size of the loan is equal to q , as an interest rate on loans to industry. More generally, let L denote the size of the loan and let D denote the amount the entrepreneur deposits in a bank (at zero interest) in period 1 to use or consume the next period.¹⁶

For simplicity, we assume that both entrepreneurs and lenders are risk neutral and do not discount future consumption. In particular, let the entrepreneur's payoff, U and the lenders' (aggregate) payoff, V , be given by:

$$(1) \quad U = q[\pi P - \varphi(\pi) - \pi(1+r) - (1-\pi)d] + D$$

$$(2) \quad V = q[\pi(1+r) + (1-\pi)d] - cL.$$

To begin with, we assume that all the bargaining power resides with the entrepreneur. For example, suppose that the financial sector is competitive and that lenders make zero

¹⁶ Thus, the entrepreneur could, in principle, borrow more than the cost of the project, though in fact she will never choose to do so.

expected profit. Then, formally, the problem facing the entrepreneur if she chooses a monitored loan is given by:

$$(3) \quad \text{Max}_{L,D,r,d,\pi} q(\pi P - \varphi(\pi) - \pi(1+r) - (1-\pi)d) + D \quad \text{subject to:}$$

$$(3a) \quad q(\pi(1+r) + (1-\pi)d) - c_G L \geq 0;$$

$$(3b) \quad w + L - D - q \geq 0;$$

$$(3c) \quad qP + D - (1+r)q \geq 0; \quad \text{and,}$$

$$(3d) \quad D - dq \geq 0;$$

where, $D \geq 0$, $L \geq 0$ and $\pi \in [0,1]$. Whereas the problem facing the entrepreneur if she chooses a non-monitored loan is given by:

$$(4) \quad \text{Max}_{L,D,r,d,\pi} q(\pi P - \varphi(\pi) - \pi(1+r) - (1-\pi)d) + D \quad \text{subject to:}$$

$$(4a) \quad q(\pi(1+r) + (1-\pi)d) - c_A L \geq 0;$$

conditions (3b), (3c), (3d), and,

$$(4e) \quad \pi \in \text{argmax}_{\pi \in [0,1]} q(\pi P - \varphi(\pi) - \pi(1+r) - (1-\pi)d);$$

where, $D \geq 0$, and $L \geq 0$.

Constraints (3a) and (4a) can be thought of as the lenders' zero expected profit or individual rationality conditions. Constraint (3b) ensures that there are sufficient funds to undertake the project. Constraints (3c) and (3d) ensure that the payments, $(1+r)$ and d , are feasible. The extra constraint (4e) arises because the entrepreneur cannot commit to a probability of success level at the time the debt contract is signed. Thus, the agreed success probability level must be consistent with the incentives facing the entrepreneur at the time at which the probability is chosen.

We chose this model design so as to be able to show most results using simple graphs. Appendix A provides a more formal treatment, but here we will just fix some intuitions. In both problems, at the optimum, $D = d = 0$, and, if the project is undertaken, $L = q - w$. That is, no money is deposited in the bank in order to make a repayment if the project fails, and the optimum loan is just sufficient to cover the entrepreneur's shortfall in funding the project. The intuition for this is that the cost of borrowing is assumed to be higher than the interest rate on deposits, so the entrepreneur will never borrow to deposit.

Figure 2.1(i) illustrates problems (3) and (4) for the case where the borrower is roughly indifferent between a German and an Anglo-Saxon style loan. The horizontal axis shows the probability of project success. The vertical axis shows $(1+r)$, the contract repayment terms after dividing through by the size of the project. That is, all vertical distances are per unit size of the project.

Consider first the monitored loan problem. The area under the curve $P - \varphi'(\pi)$ shows the entrepreneur's per unit expected profit net of private cost φ but gross of the cost of borrowing. This is maximized where the line crosses the horizontal axis, hence the first

best probability of success--that stipulated by a monitored loan contract--is given by the equation $P - \varphi'(\pi) = 0$. The hyperbola, $(1+r) = c_G(q-w)/q\pi$, represents the boundary of the lenders' zero profit constraint. That is, per unit expected repayments to lenders, $\pi(1+r)$, must be at least $c_G L/q$ for a monitored loan. The net per unit expected profit of an entrepreneur with a German style loan is thus shown by the area $OPa\pi_G$ minus the rectangle $O(1+r_G)b\pi_G$ or equivalently (exploiting the properties of hyperboli) by $OPa\pi_G$ minus the rectangle $O(1+r_A)\hat{b}\hat{\pi}$. A necessary condition for the project to be undertaken with a German-style loan is that the expected profit area is non-negative. Notice that this requires that $P > (1+r_G)$; thus constraint (3c) never binds in Problem (3).

We now move to the non-monitored loan problem. If the project fails, the entrepreneur makes no payment to the lender whereas she makes a non-zero payment, $(1+r)$, if the project succeeds. Therefore, the entrepreneur does not receive the full benefit of her actions at the margin and will only act to increase the probability of success up to the point where $P - \varphi'(\pi) = (1+r)$. The downward sloping curve in Figure 2.1(i) shows all combinations of $(1+r)$ and π that satisfy this incentive compatibility constraint. The boundary of the lender zero profit constraint is given by the hyperbola, $(1+r) = c_A(q-w)/q\pi$. Combining these two constraints gives the equation $P - \varphi'(\pi) = c_A(q-w)/q\pi$ that identifies points at which the two lines cross. Where, as shown, there are two solutions to this equation, the entrepreneur makes a higher profit at a point like a than at a point like \hat{a} in the diagram. For low values of P or high values of c_A , there will not be a solution; that is, no pair $(\pi, (1+r))$ will satisfy both constraints. In these cases, non-monitored loan contracts are infeasible. Notice that, if such a contract is feasible, $P > (1+r_A)$, so constraint (3c) never binds in Problem (4) either.

The per unit net expected profit of an entrepreneur with an Anglo-Saxon style loan is shown by the area $OPa\pi_A$ minus the rectangle $O(1+r_A)a\pi_A$. We can think of the area $\pi_A a\pi_G$ as representing the per unit dead-weight loss, **DWL**, that arises from the information asymmetry. To decide whether an Anglo-Saxon or German loan is more profitable, an entrepreneur compares the per unit dead-weight loss represented by the area $\pi_A a\pi_G$ (shaded downward) with the per unit cost difference represented by the rectangle $\pi_A \hat{a}\hat{\pi}$ (shaded upward). These are drawn to be roughly equal. If the per unit dead-weight loss from not monitoring were larger than the per unit added cost of monitoring then the entrepreneur would choose a German style loan.

For purposes of formality only, the above discussion is summarised in the following condition which we state without proof. We assume that, where feasible, an entrepreneur always chooses an Anglo-Saxon over a German style loan if indifferent and chooses to undertake the project at zero profit.

Condition 2.1: The entrepreneur will choose an Anglo-Saxon style, non-monitored loan if and only if there exists a $\pi_A \in (0,1)$ given by

$$(i) \quad P - \varphi'(\pi_A) = c_A(q-w)/q\pi_A \quad \text{and}$$

$$(ii) \quad \varphi''(\pi_A) \geq c_A(q-w)/q\pi_A^2 \quad \text{such that}$$

$$(iii) \quad \text{DWL}(P, c_A, c_G) := \left(\pi_G P - \varphi(\pi_G) \right) - \left(\pi_A P - \varphi(\pi_A) \right) \leq (c_G - c_A)(q-w)/q$$

where $\pi_G \in (0,1)$ is given by

$$(iv) \quad P - \varphi'(\pi_G) = 0.$$

Otherwise, the entrepreneur chooses a German style monitored loan, if and only if

$$(v) \quad \pi_G P - \varphi(\pi_G) - c_G(q-w)/q \geq 0.$$

Condition (i) is the feasibility condition for a non-monitored loan. Condition (ii) selects points like a over points like \hat{a} . Condition (iv) defines the first best success probability level. Condition (v) is the zero profit condition for a monitored loan. And Condition (iii) states that an Anglo-Saxon loan is preferred if the deadweight loss is less than the additional cost of monitoring. Notice that both the dead-weight loss, **DWL**, and the efficient success probability level, π_G , are well defined regardless of whether or not a solution exists to Problem (3) that yields non-negative profit. This makes it easier to formalize comparative static results.

As dead-weight losses are positive, a necessary condition for an entrepreneur to choose an Anglo-Saxon style loan over a monitored one is that it have lower cost; that is, $c_G > c_A$. In general, it would be wrong to judge the relative merits of the German and Anglo-Saxon systems simply by comparing the observed interest rate charged on loans to industry, since such charges ignore risk premia and the costs of other services provided by the lender alongside a loan.¹⁷ However, in our simple model, if solutions to both Problems (3) and (4) exist, then a necessary condition for the entrepreneur to prefer a German style loan is $r_G < r_A$. That is, *ceteris paribus*, the German system of monitored loans should only be chosen by entrepreneurs in this simple model where it results in lower interest rates to industry, r . The intuition is that π_G is always at least as large as π_A . If $r_A = r_G$ as shown in Figure 2.1(ii), then $c_G = \pi_G(1+r_A)$. In this case, the extra cost rectangle, $\pi_A ab' \pi_G$, must be larger than the deadweight loss area $\pi_A a \pi_G$.

2.1.2 Simple comparative statics of the basic model; (a) Costs of credit. Consider again levels of P , c_G and c_A such that a probability of success level π_A exists that satisfies equation 2.1(i) and that satisfies Condition 2.1(ii) with strict inequality. If both costs of providing credit, c_A and c_G , were to increase by the same amount (leaving the

¹⁷ Though this comparison is often used in the contemporary debate.

right side of Condition 2.1(iii) unchanged), then the dead-weight loss of a non-monitored loan, $DWL(P, c_A, c_G)$ would be (locally) increased.

That is, given the additional cost of monitoring, *ceteris paribus*, the higher is the cost of credit the more likely is an entrepreneur to choose a German style monitored loan. The intuition for this is that raising the marginal cost of a monitored loan, c_G , has no effect on the first-best success probability level π_G . But, raising the marginal cost of a non-monitored loan, c_A , reduces the success probability level, π_A , induced by the second-best optimal contract, raising the dead-weight loss. In Figure 2.1(iii), raising c_A to c'_A reduces expected revenues (net of private cost, φ) of the entrepreneur with an Anglo-Saxon style loan by the almost rectangular area $\pi'_A a' a \pi_A$. But raising c_G to c'_G has no effect on the expected revenues of the entrepreneur with a German style loan.

One thing that might similarly affect both the cost of providing monitored and non-monitored loans would be the cost of capital to the financial sector itself. A possible empirical proxy for this could be the real rate of interest on government debt. Figure 2.1(iv) therefore compares smoothed nominal yields on British consols with such yields on Bavarian and Prussian state bonds from 1820 to 1869 and on an index of German bonds from 1870 to 1914.¹⁸ Two things stand out. First, consol rates were lower in the late nineteenth than they were at the height of the English industrial revolution a century earlier. Since the two series move together, it is reasonable to suppose that German nominal rates also fell over this century. The fall in nominal interest rates partly reflects price movements. The late eighteenth century saw great price instability driven by harvests and wars, while the late nineteenth century is traditionally thought of as a deflationary period.¹⁹ It is likely, however, that real interest rates were also lower in the later period.

Secondly, German interest rates were consistently between a half and a whole point above their English equivalents. This could reflect higher risk premiums on less secure governments but it is also consistent with greater capital scarcity in Germany. The second observation counters the effect of the first: the expansion of heavy industry in Germany took place at a time when interest rates were low by historical standards but in a place where interest rates were high by international standards.

¹⁸ The series used were: Prussian 4s, 1820-43; Prussian 3.5s 1844-69; Bavarian 4s, 1820-40 and 1860-9; Bavarian 3.5s, 1842-59. Years where one series was missing (such as 1841) were omitted. The index of German bonds yields is that given by Homer. Homer's British series for this period is somewhere between Mitchell's and Harley's.

¹⁹ Late nineteenth century price indices are dragged down by falling food prices driven by imports from the New World. Food figures very heavily in the consumer price indices of the time. Industrial goods prices may not have fallen so sharply.

2.1.2 Simple comparative statics of the basic model; (b) Bargaining power. So far, we have assumed that the all the bargaining power lies with the entrepreneur. Next, we consider the opposite extreme: that is, where all the bargaining power lies with the lenders. In this case, the problem faced by lenders who choose a monitored loan is given by:

$$(5) \quad \text{Max}_{L, D, r, d, \pi} q(\pi(1+r) + (1-\pi)d) - c_G L \quad \text{subject to}$$

$$(5a) \quad q(\pi P - \varphi(\pi) - \pi(1+r) - (1-\pi)d) + D \geq 0.$$

and to constraints (3b), (3c) and (3d), where $D \geq 0$, $L \geq 0$ and $\pi \in [0,1]$. If the lenders chose a non-monitored loan then the problem becomes

$$(6) \quad \text{Max}_{L, D, r, d, \pi} q(\pi(1+r) + (1-\pi)d) - c_A L \quad \text{subject to}$$

constraints (5a), (3b) (3c), (3d) and (4e), where $D \geq 0$, $L \geq 0$ and $\pi \in [0,1]$.

Condition (5a) is the zero profit constraint for the entrepreneurs. The other constraints have the same interpretation as before.

Let (π_G^L, r_G^L) denote the solution to Problem (5) and let (π_A^L, r_A^L) denote the solution to Problem (6). For completeness only, we state formally but without proof, the conditions under which the lenders choose Anglo-Saxon or German style loans. As before, we assume that the lenders choose non-monitored over monitored loans if they are indifferent and that they choose to undertake the project at zero profit. The intuition is similar to that for Condition 2.1.

Condition 2.2: *The lenders will choose an Anglo-Saxon style, non-monitored loan if and only if*

$$(i) \quad \pi_G^L(1+r_G^L) - \pi_A^L(1+r_A^L) \leq (c_G - c_A)(q-w)/q \quad \text{and,}$$

$$(ii) \quad \pi_A^L(1+r_A^L) - c_A(q-w)/q \geq 0.$$

Otherwise, the lenders choose a German style, monitored loan if and only if

$$(iii) \quad \pi_G^L(1+r_G^L) - c_G(q-w)/q \geq 0.$$

The left side of Condition 2.2(i) is the difference in lenders' per unit expected revenues across loan types. The right side is the same as the right side of Condition 2.1(iii): the difference in per unit costs. Conditions 2.2(ii) and 2.2(iii) are both zero expected profit conditions. The similarity of Condition 2.2(i) to Condition 2.1(iii) enables us to compare formally the model where lenders have the bargaining power to our benchmark case where it lay with borrowers.

Consider levels of P , c_G and c_A such that a success probability level π_A exists that satisfies Conditions 2.1(i) and 2.1(ii). Then the left side of Condition 2.1(iii), the deadweight-loss that arises from the non-monitored loan where entrepreneurs have all the bargaining power, is (weakly) less than the left side of Condition 2.2(i), the difference in per unit expected revenues where lenders have all bargaining power. That is,

$DWL(P, c_A, c_G) \leq \pi_G^L(1+r_G^L) - \pi_A^L(1+r_A^L)$. This is proved formally in the appendix.

Loosely speaking, if the bargaining power is moved from entrepreneurs to the suppliers of capital then, *ceteris paribus*, we are more likely to see a German financial system emerge. One reason is that lenders attempt to extract surplus by charging a higher interest rate, r . This induces entrepreneurs with unmonitored loans to deliver lower success probability levels, but has no such effect where loans are monitored. A second reason is that lenders are unable to extract all the surplus with a non-monitored loan contract: as with efficiency wage contracts, entrepreneurs need to be given some surplus in order to induce "effort."

Again, we can see this result in a picture. In Figure 2.1(v), equal levels of lenders' per unit expected revenues are shown by hyperboli, H . Given all the bargaining power, lenders' expected revenues with a German style loan are only constrained by the entrepreneurs' zero expected profit Condition, (5a). Thus, the maximum per unit expected revenue, $\pi_G^L(1+r_G^L)$, is that represented by the rectangle, $O(1+r_G^L)b^L e_G$, and is equal to the per unit expected revenues of entrepreneurs (net of private cost, φ), $OPa\pi_G$. That is, $\pi_G^L(1+r_G^L) = \pi_G P - \varphi(\pi_G)$.

A lender's expected revenues with an Anglo-Saxon style loan are constrained by the incentive compatibility Condition (4e). Thus the maximum expected revenue is that represented by the hyperbola, H_A^L , tangent to the constraint line $P - \varphi'(\pi)$. These maximized per unit expected revenues, $\pi_A^L(1+r_A^L)$, are shown by the rectangle $O(1+r_A^L)a^L \pi_A^L$. Recall that the per unit expected revenues of entrepreneurs (net of private cost, φ) with an Anglo-Saxon style loan, when the entrepreneur has all the bargaining power, $\pi_A P - \varphi(\pi_A)$, was represented by the area $OPa\pi_A$. The area $O(1+r_A^L)a^L \pi_A^L$ is less than the area $OPa\pi_A$; that is, $\pi_A^L(1+r_A^L) \leq \pi_A P - \varphi(\pi_A)$. The difference is made up of the two shaded areas: the expected surplus retained by entrepreneurs (shaded upward) and the extra dead-weight loss due to lower probability of success (shaded downward).

It is not clear whether the balance of bargaining power lay with borrowers or lenders during the English industrial revolution, but most historians agree with Gerschenkron that early in the 1870s and 1880s in Germany, banks "acquired a formidable degree of ascendancy over industrial enterprises." Gerschenkron goes on to describe this as a "master-servant" relationship.²⁰ Banks were often able to force firms to deal exclusively with one lender, limiting competition.²¹ By the end of the century, mergers and cartelisation had shifted the balance of power back toward borrowers, but this effect was

²⁰ Gerschenkron (1962) p. 14 and p. 21. See also especially Hilferding (1910). For a dissenting view see Edwards and Ogilvie (1995).

²¹ See Da Rin (1993), pp. 22-3.

lessened by amalgamations within the financial sector itself, gradually eliminating competition from smaller, private banks.²² By 1913, 17 of the 25 largest enterprises in Germany (measured by paid up capital) were banks.²³

In addition, from 1884 onward, the state began to impose taxes on transfers of securities. This probably had two effects relevant for our purposes. First, by hindering competition, it acted "to monopolize the banking business in the hands of a few powerful concerns."²⁴ Second, it directly raised the cost of Anglo-Saxon style finance in Germany.

2.2 Extending the Model to Endogenous costs

In this subsection, we add some structure to our simple model. In particular, we model the costs of monitored and non-monitored loans as functions of firm size and of the size of the secondary asset market. Our main interest is to ask which systems of finance can emerge in an economy, taking as given (for now) both the size and number of industrial firms. For heuristic purposes, we assume that all industrial projects are the same size and that there is one project per firm. Hereafter, then, we will use the term firm and project interchangeably. Underlying the results is our basic model with all the bargaining power assigned to borrowers. We compare our results with the structure of early industry in England and Germany. We then consider two comparative statics experiments: first, increasing the size of the secondary asset market by including other forms of traded assets, such as government or merchant stock; and second, increasing the initial private wealth of entrepreneurs. Finally, we consider the welfare properties of both types of financial system.

2.2.1. Internal and external economies in debt costs. We want to consider what determines the costs of monitoring versus the cost of raising capital through non-monitored, tradeable debt. Following Diamond (1984), we will assume throughout that it is impossible to provide debt that is both monitored and tradeable. This is a reasonable first approximation since, if debt is widely traded, there are free rider problems in

²² For shifts in the balance of power, see Gerschenkron, (1962) p. 21, and Tilly (1992) See Riesser (1911) and Da Rin (1993) for details on increased concentration in banking. Riesser, himself a member of a great bank, denied that concentration reduced competition (except insofar as it squeezed out smaller concerns). But then, he also denied that cartels reduced competition in industry.

²³ Tilly (1986), pp. 113-4.

²⁴ Quoted by Riesser (1910) from Frankfurter Zeitung, May 26, 1884. Riesser has a detailed discussion of the effect of this legislation on bank concentration. See also, Tilly (1986) pp. 125-7.

monitoring.²⁵ More restrictively, we maintain our assumption that a firm can raise capital in either a monitored or a non-monitored form, but not both. In the real world, a lender may not need to provide all of a firm's external capital to have sufficient incentive to monitor. Thus, a monitored firm might choose non-monitored forms of debt for the rest of its capital. For simplicity, however, we abstract from issues of optimal capital structure completely by making it a simple zero-one choice.

Beyond these, the two key assumptions that drive the results that follow are, first, that there are scale economies in monitoring and, second, that there are externalities in the use of asset markets. That is, there are internal economies in one and external economies in the other.

It is plausible that there were scale economies in monitoring.²⁶ The cost of monitoring one large loan was probably less than the cost of monitoring two small loans. For concreteness, we assume below that these scale economies derive from fixed costs but this is not essential. What is essential, however, is that the scale economies in monitoring are assumed to be larger than any scale economies in issuing tradeable assets. This would be a dubious assumption were we to consider public issues of equity or debentures on the London stock market. There appear always to have been considerable fixed costs in such issues. However, with the notable exception of the railways, the industries of England's industrial revolution almost never raised capital directly from London's stock market.²⁷ Rather, they issued notes or bills of exchange locally that were then traded further afield. Any internal scale economies in such forms of debt were probably minor and, for simplicity, we assume that they did not exist.

The advantage of issuing tradeable forms of debt is that they provide a liquid security for lenders. If lenders care about liquidity, capital will be cheaper for borrowers if they can issue liquid forms of debt. A major reason why there are external economies in tradeable debt is that securities tend to command much lower resale or

²⁵ In modern times there are sophisticated secondary markets for semi-monitored loans such as mortgages, but we abstract from these cases. Monitoring intermediaries such as banks could themselves have tradeable liabilities but, then as now, the major liabilities of banks were non-tradeable deposits.

²⁶ Tim Guinnane (1993) has used the idea that there are scale economies in monitoring to examine the policies of German agricultural credit cooperatives in the late 19th century.

²⁷ See, for example, Mirowski (1981). One reason for this was that legislation such as the Bubble Act made it difficult to issue shares. See, Patterson and Reiffen (1990). But it is likely that the fixed costs of a London floatation would anyway have discouraged most textile firms. In 1721, there were only 8 shares regularly quoted on the London exchange and 3 of these were essentially companies for holding government debt. Until the Railways raised capital directly on London; see Neal (1995). However, the peaks in railway investment come after the high period of industrialisation in England and before our period of interest in Germany.

discount prices if secondary asset markets are thin. For concreteness, we suppose that there are fixed costs in establishing a market for a particular type of asset. These might include maintaining trading institutions such as discount houses and trading locations such as an exchange. Similarly, there are fixed costs in establishing information and communications channels for a particular asset, such as collecting and listing prices regularly in financial broad sheets. We assume that these fixed costs were divided equally over all tradeable assets in the market. Finally, tradeable debt will be cheaper the greater the variety of assets traded on the capital market because there will be greater opportunities for lenders to diversify.

2.2.2. Specific cost functions. Regardless of the existence of monitors or of a secondary asset market, we assume that it is always possible for a lender to issue a non-monitored, non-tradeable loan. We think of these as ordinary bank loans. Such debt is assumed to have constant marginal cost, v , where $v > 1$. Thus the cost of a non-monitored, non-tradeable loan of size $(q-w)$ is $v(q-w)$.

Let \tilde{N} represent the number of securities traded on the secondary asset market and let s represent the average amount per security traded, so that $\tilde{N}s$ is the total volume of capital in the market.²⁸ We assume that the marginal cost of a tradeable non-monitored loan is given by: $t(\tilde{N},s) = l + F/\tilde{N}s + f(\tilde{N})$ where $l \in (1,v)$ represents a lower bound on the cost of tradeable debt; $F > 0$, represents the fixed costs of the market; and $f(\cdot)$, where $f' < 0$, represents the gain from having more diverse assets traded on the capital market. If the existing secondary market size is large in relation to the size of an individual firm, then the entrepreneur faces an approximately constant marginal cost of tradeable debt. Thus, the cost of a tradeable non-monitored loan of size $(q-w)$ is $t(\tilde{N},s)(q-w)$.

The cost of a monitored loan of size $(q-w)$ is given by $M+mq+v(q-w)$, where $M > 0$ is the fixed cost of monitoring and $m \geq 0$ is the marginal cost of monitoring.

2.2.3 Equilibria at different numbers and sizes of firms. Figure 2.2(i) illustrates the different possible equilibria as we vary the number and size of industrial firms.²⁹ The vertical axis represents the number of industrial firms, N , in the economy. The horizontal axis represents the size, q , of these firms. Recall that we have assumed that all firms are the same size and that, for now, both size and number are given exogenously. Figure 2.2(i) is drawn for the case where entrepreneurs have no initial private wealth

²⁸ A security could be debt from one of the projects under consideration or it could be a liability from some other part of the economy that utilizes this secondary asset market.

²⁹ Again, a more formal treatment is in Appendix A.

(that is, $w = 0$) and where the number of different securities traded in the secondary asset market is bounded by the number of firms (that is, $\tilde{N} \leq N$).

There are three (or, depending on how you count, four) regions as shown in Figure 2.2(i). There is an individual firm size, $\hat{q} > 0$, such that monitored loans are strictly preferred to non-monitored, non-tradeable loans, if and only if $q > \hat{q}$. And for every uniform firm size, q , there is a number of firms, $N(q)$, below which it is not possible to sustain a secondary asset market even if all the firms issued tradeable debt.³⁰

Loosely speaking, when there are relatively few, small firms, the unique equilibrium is for all the firms to choose non-monitored, non-traded loans of the kind associated, for example, with Anglo-Saxon banks. The intuition is first that firms are too small to warrant the fixed costs of monitoring, even after allowing for deadweight losses. Second, even if all firms were to choose tradeable debt, the secondary asset market would still be too small for such debt to be cheaper than bank loans.

When there are relatively few, large firms, the unique equilibrium is for all the entrepreneurs to choose monitored loans of the type associated with German banks. The costs of monitoring are now spread thinly enough to warrant monitoring in order to avoid the dead-weight losses associated with Anglo-Saxon loan contracts. Moreover, even if all firms issued tradeable, non-monitored debt, its cost plus associated deadweight losses would outweigh the costs of monitoring.

When there are relatively many firms of medium size, there is an equilibrium in which all entrepreneurs choose tradeable non-monitored loans. In this case, the size of the secondary asset market is such that tradeable debt costs less than non-monitored bank loans. Moreover, the cost saving from not monitoring is enough to overcome the associated deadweight losses. But this is not the only equilibrium in this region. For example, if all entrepreneurs were to choose non-monitored bank loans then no entrepreneur would want to issue tradeable debt since it would have to bear the full fixed costs of the market on its own. Thus, if the size of firms is small, there is also an equilibrium in which all entrepreneurs choose non-monitored non-tradeable loans. Similarly, if the size of firms is large, there is an equilibrium in which all entrepreneurs choose monitored non-tradeable loans as, again, the fixed costs of trading debt are too large for it to be worthwhile for just one firm to do.

There are still other equilibria in this region. For example, if $q < \hat{q}$, we can

³⁰ The exact shape of $N(q)$ above \hat{q} need not be as shown. Loosely speaking, it will always be upward sloping if the fixed costs of monitoring, M , greatly exceeds the fixed costs of the secondary market, F , spread across $N(\hat{q})$ firms. That is, $M \gg F/N(\hat{q})$. More generally, even if it is downward sloping to the right of \hat{q} , the function $N(q)$ will always be cut from above by hyperboli.

construct an equilibrium in which just enough entrepreneurs choose tradeable, non-monitored loans for each entrepreneur to be indifferent between these and non-tradeable, non-monitored loans. If we ignore genuine mixed strategy equilibria, there will typically be one such 'split' equilibrium for each realisation of N and q . If $q = \hat{q}$, then there is a continuum of equilibria since entrepreneurs are indifferent between monitored and non-monitored, non-tradeable loans. Otherwise, if $N = N(q)$ there is no equilibrium in which only some entrepreneurs choose the market.

2.2.4 The Size and Number of firms in the British and German industrial revolutions.

How does this simple picture compare to the experience of the British and German industrial revolutions? In Gerschenkron's famous comparison of early and late industrialisers, alongside his observations on the different roles of banks, he also noted that "the more backward a country's economy, the more pronounced was the stress in its industrialisation on bigness of both plant and enterprise."³¹ The size and capital requirements of the typical industrial firms in late 18th century Britain were quite small. Landes, for example, writes:

"The early machines, complicated though they were to contemporaries, were nevertheless modest, rudimentary, wooden contrivances which could be built for surprisingly small sums. A forty-spindle jenny cost perhaps £6 in 1792; The only really costly items of fixed investment in this period were buildings and power, but here the historian must remember that the large, many-storeyed mill that awed contemporaries was the exception. Most so-called factories were no more than glorified workshops: a dozen workers or less; one or two jennies, perhaps, or mules; and a carding machine to prepare the rovings."³²

As late as 1841, Gatrell found that the median Lancashire cotton primary processing firm had just over 100 employees while the median firm in subsidiary textile production had fewer than 50 employees. The same study identified over 1000 separate firms in the cotton textile industry in Lancashire alone.³³

The size of the typical German firm by the 1870s was much larger. There were two underlying reasons for this. First, industrial firm sizes generally increased after 1850. Even in the British textile industry the number of spindles per firm in the British spinning industry increased by 50% from 1850 to 1870. But in Germany, it increased by 600%.³⁴ Second, whereas textiles were the "leading sector" of the first industrial revolution, the second industrial revolution was dominated by heavy industries such as steel and chemicals with larger fixed capital requirements and larger optimal plant sizes.

³¹ Gerschenkron (1962) p. 354.

³² Landes (1969) pp.64-5. See also Crouzet (1963) pp.164-75; and Pollard (1964) pp.301-5.

³³ Gatrell (1977), p. 98.

³⁴ Landes (1969) pp. 223-4.

Very loosely speaking, Germany industrialised at a time when firms were larger and industrialised into sectors dominated by larger firms. Moreover, both these effects seem to have been more pronounced in Germany than elsewhere. In steel smelting, for example, by the turn of the century, the median member of the German steel cartel was four times bigger than its equivalent firm in Britain.³⁵

Within Germany, the *Grossbanken* appear to have favoured larger firms. As early as 1853, the stated policy of the Bank of Darmstadt was to concentrate on firms with a turnover of 50,000 Guilders.³⁶ Tilly has called this "Development Assistance for the strong",³⁷ while Gerschenkron argued that this later resulted in a sectoral bias:

"until the outbreak of World War I, it was essentially coal mining, iron and steel making, electrical and general engineering, and heavy chemical output which became the sphere of activity of German banks. The textile industry, the leather industry, and the foodstuff producing industries remained on the fringes of bank interest ... it was heavy rather than light industry to which the attention was devoted."³⁸

One way to contrast industrial revolutions is to compare different economies at the same given stage of development as measured by the size of their industrial sector.³⁹ Figure 2.2(ii) represents the volume of capital invested in an industrial sector by hyperbolic superimposed on the previous diagram. A reasonable summary of the stylised facts would be that the late 18th century British economy was represented by an area like *A* in the figure: with relatively many, relatively small industrial firms. Late 19th century Germany looked more like the area *G* in the figure. This concurs with our simple model. In Britain, external finance was generally unmonitored, sometimes from small local banks and sometimes in tradeable forms such as bills of exchange.⁴⁰ In Germany, though we should not exaggerate its role, industrial finance provided by large banks was important.

2.2.5. Simple comparative statics: (a) Other Traded Securities. If the secondary asset market includes securities other than those issued by industrial companies, such as government debt instruments and merchant paper, then this relaxes the constraint that

³⁵ Ibid. p.263.

³⁶ Quoted by Tilly (1986), p. 121. Landes (1969, p. 208) argues that the *Grossbanken* "sought out the largest possible clientele."

³⁷ Ibid. See also Tilly and Fremling (1976) esp. p. 420; and Tilly (1982).

³⁸ Gerschenkron (1962) p. 10. See also, Tilly (1986), esp. p.148: "the creditbanks concentrated on large enterprises." Neuburgher and Stokes (1974) have argued that this bias may have actually slowed German growth.

³⁹ Cf. Crafts (1985) compares economies at given levels of per capita income.

⁴⁰ For the use of tradeable bills of exchange to finance early British industrial investment especially in the Lancashire textile regions, see Ashton (1945 and 1955, ch.6); and Anderson (1970).

$\tilde{N} \leq N$. The effect is shown in Figure 2.2(iii). The area in which there are equilibria involving tradeable non-monitored debt is increased: its lower boundary shifts down from $N(q)$ to $\hat{N}(q)$. Intuitively, any fixed costs of the market are spread more thinly. Thus, *ceteris paribus*, an entrepreneur in an economy with an active secondary market in non-industrial debt is more likely to choose tradeable (and hence non-monitored) forms of debt.

The use of tradeable debt by industrial firms in Britain would itself have helped develop the market and encouraged others to follow. But this process was probably aided by the early existence of a secondary asset market in Britain trading in government and merchant debt. Landes writes that

"in no country in Europe in the 18th century was the financial structure so advanced and the public so habituated to paper instruments as in Britain The development of a national network of discount and payment enabled the capital hungry industrial areas to draw ... on the capital rich agricultural districts."⁴¹

The secondary market in government paper was particularly well developed, having recovered from crises in the 1720s. There is evidence consistent with integration between this market and investment in the industrial north by at least the last quarter of the century.⁴² England's eighteenth century internal capital markets were probably more developed than those of newly united Germany a century later. Our model suggests that early, well developed secondary asset markets in England may have influenced the form of loans chosen during the industrial revolution.

2.2.5. Simple comparative statics: (b) Increasing Wealth. Given the importance of self-finance in the British industrial revolution, we would like our model to be robust to changes in the initial private wealth of entrepreneurs. The effect of increasing entrepreneurs' wealth, w , are shown in Figure 2.2(iv). If $q \leq w$, entrepreneurs will entirely self-finance their projects as shown by the region on the left of the figure. If $q > w$, they will only borrow their shortfall, $(q-w)$, so, as we increase w , the amount borrowed is reduced.

This affects the trade-off between non-monitored non-tradeable debt and monitored non-tradeable debt in two ways. First, the fixed costs of monitoring are spread over a smaller loan. Second, while the efficient success probability level, π_G , is unchanged, that chosen by entrepreneurs with non-monitored loans, π_A , is increased. The intuition for this is the same as that represented in Figure 2.1(iii): since less is borrowed, less needs be repaid if the firm is successful, and the entrepreneur receives a larger share of

⁴¹ Landes (1969) p.74.

⁴² See, for example, Buchinsky and Polak (1993) and Hoppitt (1986).

the return to effort at the margin. Both these effects favour non-monitored non-tradeable over monitored non-tradeable loans, so \hat{q} is unambiguously shifted to the right in Figure 2.2(iv).

Next consider the effect of increasing private wealth, w , on tradeable non-monitored loans. In addition to the above effects favouring not monitoring, there are two additional effects. First, if the wealth of all entrepreneurs is increased simultaneously, then the volume of assets in the secondary market will be reduced as borrowing is decreased. This will increase the cost of tradeable loans. Second, suppose that tradeable loans are strictly preferred to non-monitored, non-tradeable loans, but are indifferent to monitored, non-tradeable loans (as occurs along the border $N(q)$ when $q > \hat{q}$). Then $t < v$. That is, the marginal saving, t , to a tradeable, non-monitored loan as personal wealth increases is smaller than the marginal saving, v , to a monitored, non-tradeable loan. From the first of these extra effects, it follows that increasing personal wealth favours non-tradeable non-monitored loans over tradeable non-monitored loans. But, we cannot say whether or not increasing wealth favours tradeable, non-monitored loans over non-tradeable, monitored loans. Here the effects conflict. This is illustrated in Figure 2.2(iv).

Overall, the main effect of increasing wealth is to increase the proportion of self-finance, and to shift the non-monitored non-traded loan region to the right. Under plausible assumptions, the border between the other two regions twists as shown. Then, at least for lower N , increasing the initial wealth of entrepreneurs generally raises the minimum size at which monitored loans are chosen.

This picture is roughly consistent with the British experience. To use Landes as our authority again, "a good many of the early mill owners were men of substance", often with wealth built up from merchant activities, putting out or even artisanal production within the sectors they later revolutionised.⁴³ For example, more than half the cotton spinning mills established in the Midlands from 1769 to 1800, were set up by people already established in some part of the textile industry.⁴⁴ Not only was the scale of the new industries small, but "18th century Britain enjoyed ... more wealth and income per head than the unindustrialised countries of today."⁴⁵ Thus, it is not surprising that Crouzet's classic study of how industrialization was funded in Britain concluded that the "simple answer to this question is the overwhelming predominance of self-finance." Germany's industrialisation, though later, started from a lower base in terms of accumulated

⁴³ Landes (1969) pp. 65-6. See also Pollard (1964) p. 300; and Crouzet (1963).

⁴⁴ Ibid. If you include the 15 mills established by Arkwright, the share is almost 3/4.

⁴⁵ Ibid. p. 78.

industrial wealth. As the model predicts, early industrial entrepreneurs in Britain borrowed less overall and their relatively small external capital requirements were met by local bank loans, promissory notes and bills of exchange. The larger requirements of their counterparts in Germany were met, at least in part, by the direct involvement of the industrial banks.

It may be that as wealth increased our stylized British industrial sector was pushed into the region where non-monitored bank loans predominate over tradeable loans. Interestingly, the high period of (local) bank loans to industry in Britain appears to have been the middle decades of the 19th century, after a generation of industrial growth and accumulation⁴⁶ Increasing personal wealth may also help explain why Britain did not converge to a German style financial system as the size of her firms increased in the 19th century. Since \hat{q} is increased with w , the British economy was less likely to be pushed into the region where monitored loans form part of the equilibria.

2.2.6. Welfare and Policy.

Our model may have implications beyond the British and German industrial revolutions. Where parameters are such that there is only one equilibrium, it is plausible that a suitable financial system will emerge on its own. However, where there are multiple equilibria there may be scope for government intervention. In the model, whenever an equilibrium exists in which all entrepreneurs choose tradeable non-monitored debt, it is the Pareto dominant equilibrium. The inefficiency of the other equilibria arises from a coordination failure: if they all chose tradeable debt then the fixed costs of the market would be spread thinner and the market would be more diversified. There is an externality: no single entrepreneur takes into account her effect on the market costs of others when she makes her private decision.

It is interesting to note that the 1933 Glass-Steagall Act outlawed universal banking in the US, thus pushing the economy away from a German style finance system. Our model suggests that one argument that could be used by opponents of universal banking was that it inhibited the growth of secondary capital markets. It is possible that the same arguments would be relevant in Eastern Europe today. If it would be preferable to move toward an equilibrium with active secondary asset markets--that is, if in the long-run there will be many small firms--there may at least be a case for promoting such markets. A further reason for subsidising the use of tradeable debt is the externality associated with self-finance. The example of 18th century Britain, suggests that one way to do this

⁴⁶ See, for example, Collins and Hudson (1979); and Collins (1990). Collins and Hudson (p. 78) found that, although there were personal links between local banks and industry, banks were not closely involved with management unless a firm defaulted.

is for the government to issue debt in the form of transferable securities as shown in Figure 2.2(iii).

3: ENTRY, PERSISTENCE AND WELFARE

In the previous section, we examined the choice by individual firms of German or Anglo-Saxon style loans taking as exogenous both the size and the number of projects to be funded. In this section, we extend the model to allow both the size and number of industrial firms and the financial system to be endogenous. The idea we want to capture is that lenders may take the organisation of firms as given when they choose what type of financial institutions to form. Similarly, industrial entrepreneurs may take financial institutions as given when they organise firms. An economy may then find itself with German style banks and German style firms, each designed taking the other as given; or with Anglo-Saxon style financial institutions and Anglo-Saxon style firms. In the model below, we loosely describe these as *G* and *A*-types of equilibrium. Such a model may help explain how it is possible for both types of equilibrium to persist in neighbouring economies; that is, why did the financial systems of Germany and England not converge sooner.⁴⁷

Given the current and historical debates over whether German or Anglo-Saxon financial systems are better, it is of interest to compare the two types of equilibrium in welfare terms. We show that, under certain parameter values, it is possible for both types of equilibrium to be supported in our model. Multiple equilibria leave scope for coordination failures: that is, equilibria may be Pareto ranked. We show that for some parameter values for which both types of equilibrium can be supported, the *A*-type is Pareto dominant, but that this is not always the case. Indeed, sometimes the *G*-type of equilibrium may result in greater social surplus.

In the model here, we use project scale as our example of an organisational choice of firms, but the general idea could be extended. In practice, the organisation of firms could refer to types of machinery or to administrative forms. There are many reasons why external monitoring may be costly given the set up of a particular firm. For example, firms set up to be run by family members who have special knowledge of the business may not easily be adapted to allow for external monitoring. In general, the organisational structure of firms is more likely to be an obstacle to switching to monitored finance than switching the other way. This is consistent with our model.

Similarly, one could imagine many different financial institutional forms. But for

⁴⁷ Notice, however, that what follows is a one-shot model. What we have in mind is that each player takes the institutions established by others as given before making their own choices. We use equilibrium in a one-shot game as a metaphor for this idea.

our purposes, we restrict attention to those institutions that can monitor loans and those that can deal in tradeable assets. We call the former industrial banks and the latter discount brokers. As before, we assume that loans are either monitored or traded (or neither) but not both. We assume that discount brokers are not equipped to monitor industrial loans, but that both they and industrial banks can issue non-tradeable, non-monitored loans. As before, for simplicity only, we assume that firms must choose either to have all monitored or all non-monitored finance.

3.1 The Entry Model:

3.1.1 Costs. To allow choice of scale to be endogenous, we introduce explicit cost functions. Specifically, let $k(q)$ be the capital required to fund a project of size q , where $k, k', k'' > 0$ for all $q > 0$. Assume that average capital requirements are U-shaped; that is, there exists a unique $q^* > 0$ that minimizes $k(q)/q$. To keep things simple, assume that no potential entrepreneurs have any private wealth so the entire capital for a project must be borrowed. Also assume that there are no other traded assets except the traded debt of the projects concerned.

If a project of size q is funded by a monitored loan then the average cost of the loan are $(vk(q)+mq+M)/q$. Since, $k'' > 0$ and there are fixed costs to monitoring, the efficient scale for a project funded by such a loan, q_G^* , is greater than q^* . The average capital cost of a non-monitored, non-traded loan is just $vk(q)/q$ so the efficient project scale given such a loan q_B^* equals q^* . Again for simplicity, we assume that the marginal capital cost of a tradeable loan can only take two values; high when the asset market is thin, low when the asset market is thick. Formally, if the set of firms issuing tradeable assets is A and if the sizes of those firms are given by the vector $(q_i)_{i \in A}$, then let

$$t(A, (q_i)_{i \in A}) = \begin{cases} \hat{t} & \text{if } \sum_{i \in A} k(q_i) < V \\ t & \text{otherwise} \end{cases}$$

be the (per dollar) marginal cost of a tradeable loan, where $\hat{t} > v > t$. Assume that V and \hat{t} are sufficiently large that it is never profitable for one firm to shoulder the entire costs of the market. The total cost of a tradeable loan of size q when the market is thick is then $tk(q)$ and the efficient project size given such a loan q_A^* equals q^* .

3.1.2 Supplies and Demands. To allow the number of projects to be endogenous, we introduce entry. Assume that there are a large number of potential entrepreneurs each with an identical project, each producing the same product. There are an even larger number of potential lenders and the supply of capital is infinitely elastic. The probability that a project succeeds, given that it is funded, depends only on the actions of that project's entrepreneur. That is, project risks are assumed to be independent.

Final inverse demand for the product is given by $p(Q) = \alpha - \beta Q$ where p is price and Q

is total output.⁴⁸ A highly stylised futures market is organised as follows. An auctioneer announces a futures price, P , and active firms sell their entire contingent product, q_i , to the auctioneer before uncertainty is resolved. That is, firms do not speculate on their own account. The auctioneer aims to minimize the difference between her announced price, P , and the expected final output price, $p(Q)$, which is a function of the number and sizes of active firms and their success probabilities. Futures contracts cannot be breached: if a project succeeds, the firm must supply q_i to the auctioneer who pays whichever is greater of the announced price P or zero. We assume that the auctioneer has deep pockets. Our stylised futures market is a device to ensure that the entrepreneur receives the expected price (not the actual final price) in any equilibrium. If entrepreneurs were to receive the uncertain final price, $p(Q)$, then it would be possible even in equilibrium for so many projects to succeed as to drive the final price down such that even "successful" firms go bankrupt. Futures markets such as ours insulate producers from price risk.

3.1.3. Timing. The timing of the game is as follows. All actions within periods are simultaneous unless otherwise stated. Everything except unmonitored success probability level choices is observable between periods.

Period 0: Potential lenders choose either (A) to form discount houses able to grant and trade Anglo-Saxon type loans; or (G) to form industrial banks able to grant and monitor German style loans. Potential entrepreneurs each choose a scale for their project ($q_i \geq 0$) should it be realised. There are no sunk costs incurred.

Period 1: Potential entrepreneurs choose: (A) to offer to lenders a tradeable non-monitored loan contract; or (B) to offer a non-tradeable non-monitored loan contract, in each case specifying the loan size, $k(q_i)$ and the repayment rule $R_i \geq 0$; or (\emptyset) to "pass."⁴⁹

Period 2: Those potential entrepreneurs who passed before choose either (G) to offer to lenders a monitored loan contract, specifying the loan size $k(q_i)$ and the repayment rule $R_i \geq 0$, and success probability level ($\pi_i \in [0,1]$); or (\emptyset) to pass again.

Period 3: First, A -contracts are randomly assigned to lenders who chose A ; G -contracts to lenders who chose G ; and B contracts to both, such that each lender is offered at most one contract. Then, lenders simultaneously either accept or reject the offers. If loans are made, the relevant costs are incurred.

Period 4: The auctioneer announces a price, P . Active entrepreneurs submit

⁴⁸ That is, we allow the price to go negative at very large Q .

⁴⁹ For ease of notation, the repayment rule, R , is no longer written in the form of an interest rate on the loan. If the firm is solvent, it pays back Rq .

contingent supplies to the market; that is, they agree to supply q_i contingent on the success of their project. (Simultaneously) all active entrepreneurs with A-loans choose success probability levels ($\pi_i \in [0,1]$), while all those with G-loans implement the specified levels.

Period 5: All uncertainty is resolved: that is, for each active project, output is either q_i with probability π_i or 0 with probability $(1-\pi_i)$. For successful projects, deliveries are made to the auctioneer, who pays either P or 0 whichever is greater. Debtors pay creditors whichever is the smaller of the realised revenue of the firm, that is, $\max\{Pq, 0\}$, or the agreed loan repayment, $R_i q_i$.

Some assumptions implicit in the timing above are worth comment. The assumption of no sunk costs in period 0 avoids hold-up problems. The contracts are only allowed to be contingent on the success of the project in question. In particular, to keep things simple, contracts can not be made contingent on other firms' contracts, prices or the number of projects accepted. Monitored loan contracts are proposed after non-monitored contracts. Without this assumption, it is too easy to sustain equilibria: deviations can be deterred by flooding the market with loss making monitored firms. The assumption that each potential lender has access to at most one contract is to capture the idea that no lender has "market power."

3.1.4. Payoffs: Expected payoffs for the game as functions of strategies are given below. These appear complicated but they simplify once we focus attention on equilibrium (where, for example the auctioneer's announced price will never be negative). Entrepreneurs whose offers are rejected, lenders who accept no offers and traders who do not trade receive zero utility. An active entrepreneur with project size q , success probability level π , repayment rule R , and given futures price P has expected utility:

$$U(q, \pi, R, P) := \text{Max}\{\pi Pq - \pi Rq, 0\} - q\varphi(\pi)$$

If the loan was monitored then the lender has expected utility:

$$V_G(q, \pi, R, P) := \text{Min}\{\max\{\pi Pq, 0\}, \pi Rq\} - [M + mq + vk(q)].$$

If it was a non-monitored non-tradeable loan then the lender's expected utility is

$$V_B(q, \pi, R, P) := \text{Min}\{\max\{\pi Pq, 0\}, \pi Rq\} - vk(q).$$

If it was a tradable loan and if the set of firms issuing tradeable assets is A where the sizes of those firms are given by the vector $(q_i)_{i \in A}$, then the lender's expected utility is given by:

$$V_A(q, \pi, R, s, P; A, (q_i)_{i \in A}) := \text{Min}\{\max\{\pi Pq, 0\}, \pi Rq\} - t(A, (q_i)_{i \in A})k(q).$$

Let $p(B, (q_i)_{i \in B}, (\pi_i)_{i \in B})$ be the expected final output price given a set of active firms B of sizes (q_i) and probabilities of success (π_i) . Then the payoff to the auctioneer who announces future price P is given by:

$$V_T(P; B, (q_i)_{i \in B}, (\pi_i)_{i \in B}) := |P - p(B, (q_i)_{i \in B}, (\pi_i)_{i \in B})|.$$

Some simplifications follow from considering equilibria. First, in any equilibrium of any period 4 subgame in which there are active firms, the auctioneer will set future price P equal to the expected final prices given the set of active firms B , their sizes $(q_i)_{i \in B}$, and their (equilibrium) probabilities of success $(\pi_i)_{i \in B}$; $P = \alpha - \beta \sum_{i \in B} \pi_i q_i$. Second, both lenders and entrepreneurs must have non-negative expected profit in any (pure strategy) subgame perfect equilibrium. Therefore, the auctioneer's price P will be greater than any active firm's repayment rule R which, in turn, will be greater than zero. This greatly simplifies the payoff's above: loosely speaking it removes the need for the "max" and "min" notation.

3.2. Subgame Equilibria.

We are interested in two types of putative equilibrium to the above game. One type involves all potential lenders forming discount houses, and all potential entrepreneurs choosing scale q_A^* and offering tradeable, non-monitored loan contracts, of which some number are selected. We refer to these as putative A -equilibrium paths. The other type involves all potential lenders forming industrial banks, and all potential entrepreneurs choosing scale q_G^* and offering non-tradeable monitored loan contracts, of which again some number are selected. We refer to these as putative G -equilibrium paths.

Our aim is to give conditions under which each of these types of equilibria exist and then to compare them. We will only consider pure-strategy, subgame-perfect equilibria.⁵⁰ Hereafter, we will simply term these equilibria. We consider first the A - and then the G -equilibria in isolation, without allowing deviations in the type of loan. The forms of equilibria in these subgames turn out to be interesting in their own right.

One difficulty in formally constructing equilibria is that the number of accepted firms, N , must be an integer. This can lead to existence problems. We show below, however, that if the economy is large then equilibria ignoring the integer constraint are close to equilibria when it is imposed. Loosely speaking certain equalities in the equilibrium conditions will be relegated to approximate equalities when N is confined to be an integer.

3.2.1. Putative A -Equilibria. Our first task is to examine the putative A -equilibria abstracting from the possibility of deviating to monitored loans. Consider then the

⁵⁰ A formal note. In fact, we will only consider things close to equilibrium paths. In particular, we will not describe how an equilibrium dictates play in a subgame that can only be reached by *simultaneous* multi-person deviations. We will not even show that a pure-strategy, subgame-perfect equilibrium exists in such subgames. Thus, formally, we do not show that the equilibrium paths we describe belong to pure-strategy, subgame-perfect equilibria.

period 1 subgame following the formation of discount houses by all potential lenders and the choice of scale q_A^* by all potential entrepreneurs in period 0. Hereafter, we refer to this as the A -subgame. We are interested in putative equilibria of this subgame when the economy is large. We want the economy to be large for two reasons. First, we want there to be enough assets in the economy such that the marginal cost of using tradeable debt is lower than that of using non-tradeable debt. Second, we want there to be enough active firms in equilibrium such that the restriction that the number of firms be an integer is unimportant. The simplest way to construct a large economy in our setting is to consider cases where the demand slope parameter β is small; that is, loosely speaking such that each active firm's output has only a small effect on output prices. This is how we shall proceed below.

Recall from section 2 that if the output price is too low then no unmonitored loan contract is viable in the sense that the contract induces expected loan repayments as large as the cost of the loan. To make our model interesting therefore, we assume that, for sufficiently small demand parameter β (that is, for a sufficiently large economy), there would exist a viable non-monitored loan contract at least in the extreme case where there is only one active firm. This is achieved by Assumption A.1(a). Assumption A.1(b) is a minor technical assumption that simplifies the proofs of the propositions that follow.

Assumption A.1: (a) *There exists a probability of success $\bar{\pi}_A$ such that: $\alpha - \varphi'(\bar{\pi}_A) - vk(q_A^*)/q_A^*\bar{\pi}_A > 0$; (b) φ'' is bounded on the interval $[0, \pi_1]$ where π_1 is given by: $\alpha - \varphi'(\pi_1) - tk(q_A^*)/q_A^*\pi_1 = 0$ and $\varphi''(\pi_1) - tk(q_A^*)/q_A^*\pi_1^2 < 0$.*

We are now ready to describe putative equilibrium paths in the A -subgame. Let $[R, N, \pi, P]$ denote a path in the A -subgame along which: in period 1, all potential entrepreneurs offer the non-monitored loan contract R ; (in period 2 there are no movers); in period 3, exactly N such contracts are accepted by lenders; and, in period 4, the auctioneer announces price P and all active entrepreneurs choose probability of success π .

Proposition 3.1 describes an equilibrium path of the A -subgame when the economy is large. Specifically, it describes that equilibrium path that results in the lowest output prices. The proposition has two parts. The first part describes the path ignoring the restriction that the number of firms be an integer. The second part shows that this approximation to the true equilibrium path is "kosher" when the economy is large. That is, if we take a sequence of A -subgames for decreasing values of β then we can find a sequence of equilibrium paths of these games such that the equilibrium repayment rules, probabilities of success and output prices converge to those ignoring the integer restriction and such that the equilibrium integer number of firms is always within 1 of

that ignoring the integer restriction.

Proposition 3.1: *Given Assumption A.1, suppose that, in period 0, all potential entrepreneurs choose size q_A^* and all potential lenders choose to form discount houses.*

(a) *Consider the subsequent A-subgame ignoring the restriction that the number of active firms be an integer. Then, for sufficiently small β , the path $[R_A^*, N_A^*(\beta), \pi_A^*, P_A^*]$ (uniquely) defined by*

$$\begin{aligned} \text{(i)} \quad P_A^* &= \alpha - \beta q_A^* \pi_A^* N_A^*(\beta) & \text{(ii)} \quad P_A^* - \varphi'(\pi_A^*) &= R_A^* \\ \text{(iii)} \quad R_A^* &= tk(q_A^*)/q_A^* \pi_A^* & \text{(iv)} \quad \varphi''(\pi_A^*) &= tk(q_A^*)/q_A^* (\pi_A^*)^2. \end{aligned}$$

is an equilibrium path. Moreover, P_A^ is the lowest price that can emerge on an equilibrium path of any period 3 subgame in which there are active entrepreneurs with non-monitored loan contracts.*

(b) *Consider the subsequent A-subgames with the restriction that N be an integer defined for a sequence $(\beta_n)_{n=1}^{\infty}$ such that $\beta_n \rightarrow 0$ as $n \rightarrow \infty$. Then, for sufficiently large n (small β_n), there exists a sequence of equilibrium paths of these A-subgames, $([R_A(\beta_n), N_A(\beta_n), \pi_A(\beta_n), P_A(\beta_n)])$, such that $[R_A(\beta_n), \pi_A(\beta_n), P_A(\beta_n)] \rightarrow [R_A^*, \pi_A^*, P_A^*]$ as $n \rightarrow \infty$, and such that $N_A^*(\beta_n) - 1 < N_A(\beta_n) < N_A^*(\beta_n) + 1$.*

Proof: See Appendix 2.

A partial intuition is as follows. Equation 3.1(i) is the market clearing condition. Equation 3.1(ii) is the entrepreneurs' incentive compatibility condition. Equation 3.1(iii) is the lenders' zero profit condition. A solution to these three equations is shown in Figure 3.2(i). Equation 3.1(iv) is represented by the tangency in Figure 3.2(i) between the lines representing the borrower's incentive compatibility and the lender's zero profit constraints. Thus, Equation 3.1(iv) reflects the fact that this is the lowest equilibrium price that can coexist with active non-monitored firms: for any price lower than P_A^* , there is no intersection between these constraint lines. This also partially explains why this is an equilibrium (ignoring the integer constraint). If any "deviant" non-monitored firm is accepted in addition to the $N_A^*(\beta)$ "incumbents" then, regardless of the deviant's contract, the period 4 equilibrium price will be strictly lower than P_A^* and the expected repayment to the deviant's lender will be smaller than his loan. Loosely speaking, the lines in Figure 3.2(i) representing the entrepreneurs' incentive compatibility and the lenders' zero profit conditions would not intercept. It is "as if" the lender to the would-be entrant takes the number of other incumbents as given and thus decides to reject the loan proposal.

The proof for part (b) where we respect the integer constraint is more cumbersome,

but the idea is to approximate the equilibrium of part (a) by accepting the integer number of firms just larger or just smaller than $N_A^*(\beta)$. The candidate equilibrium is then constructed by respecting the market clearing, incentive compatibility and lender's zero profit constraints but slightly relaxing the tangency condition of Equation 3.1(iv). Intuitively, as β becomes small, the effect of the integer constraint becomes small and these candidate equilibria become very close to that described in part (a). That is we get close to the tangency of Figure 3.2(i); close to the minimum sustainable price P_A^* . Thus, it is not surprising that no "deviant" could satisfy both constraints though showing this formally takes some work.

Despite the fact that we allowed free entry, entrepreneurs make positive expected profit in this equilibrium represented, represented by the shaded area in Figure 3.2(i). This is analogous to labour market equilibria in efficiency wages models. There, excluded workers would strictly prefer to enter employment at the going wage, but potential employers choose not to hire them since the resultant slight lowering of the wage would remove the incentive to work hard enough to justify employment. Here, excluded entrepreneurs would strictly prefer to enter production at the reigning repayment rule but potential lenders do not accept such contracts since the resultant slight lowering of the expected output price would remove the incentive for entrepreneurs keep the probability of project success high enough to justify supplying the capital. We have translated an idea of an efficiency wage model to a competitive entry model with moral hazard while solving the integer problem. Although the analogy is good, notice that it is not an exact translation.

3.2.2. Putative G -equilibria. We next examine putative G -equilibria abstracting from the possibility to deviate to a non-monitored loan. Consider the period 1 subgame following the formation of investment banks by all potential lenders and the choice of scale q_G^* by all potential entrepreneurs in period 0. Hereafter, we refer to this as the G -subgame. As before, we will only consider equilibria of this subgame when the economy is large; β is small. Also as before, we need to assume that there exists a viable monitored loan contract at least in the extreme case where there is only one active firm.

Assumption A.2. *There exists a probability of success $\bar{\pi}_G$ such that $\bar{\pi}_G \alpha - \varphi(\bar{\pi}_A) - [vk(q_G^*)+M+mq_G^*]/q_G^* > 0$.*

Let $[R, N, \pi, P]$ denote a path in the G -subgame along which: (in period 1, there are no movers); in period 2, all potential entrepreneurs offer the monitored loan contract (R, π) ; in period 3, exactly N such contracts are accepted by lenders; and, in period 4, the auctioneer announces price P and all active entrepreneurs implement π . Proposition 3.2

describes an equilibrium path of the G -subgame when the economy is large. Specifically, it describes that equilibrium path that results in the lowest output prices. Like the previous proposition, the first part describes the path ignoring the restriction that the number of firms be an integer. The second part shows that this approximation to the true equilibrium path is "kosher" when the economy is large.

Proposition 3.2: *Given Assumption A.2., suppose that, in period 0, all potential entrepreneurs choose size q_G^* and that all potential lenders choose to form investment banks.*

(a) *Consider the subsequent G -subgame ignoring the integer restriction on the number of active firms. Then, for sufficiently small β the path $[R_G^*, N_G^*(\beta), \pi_G^*, P_G^*]$ (uniquely) defined by*

$$\begin{aligned} \text{(i)} \quad P_G^* &= \alpha - \beta q_G^* \pi_G^* N_G^*(\beta) & \text{(ii)} \quad \pi_G^* P_G^* - \varphi(\pi_G^*) &= R_G^* \pi_G^* \\ \text{(iii)} \quad R_G^* &= [vk(q_G^*) + M + mq_G^*] / q_G^* \pi_G^* & \text{(iv)} \quad P_G^* - \varphi'(\pi_G^*) &= 0 \end{aligned}$$

is an equilibrium path. Moreover, P_G^ is the lowest price that can emerge on an equilibrium path of any period 2 subgame in which there are active entrepreneurs with monitored loan contracts.*

(b) *Consider the subsequent G -subgame with the integer restriction on the number of active firms defined for a sequence $(\beta_n)_{n=1}^{\infty}$ such that $\beta_n \rightarrow 0$ as $n \rightarrow \infty$. Then, for sufficiently large n (small β_n), there exists a sequence of equilibrium paths of these G -subgames, $([R_G(\beta_n), N_G(\beta_n), \pi_G(\beta_n), P_G(\beta_n)])$, such that $[R_G(\beta_n), \pi_G(\beta_n), P_G(\beta_n)] \rightarrow [R_G^*, \pi_G^*, P_G^*]$ as $n \rightarrow \infty$, and such that $N_G^*(\beta_n) - 1 < N_G(\beta_n) < N_G^*(\beta_n)$.*

Proof: See Appendix 2.

A partial intuition is as follows. Paralleling the previous proposition, Equation 3.2(i) is the market clearing condition and Equation 3.2(iii) is the lenders' zero expected profit condition. Equation 3.2(ii) is the entrepreneurs' zero expected profit condition. A solution to these three equations is shown in Figure 3.2(ii). Equation 3.2(iv) represents the fact that π_G^* is the efficient success probability given the price P_G^* ; also shown in the figure. At any lower price, even efficient actions by the active entrepreneurs will not generate sufficient expected revenues to satisfy the two zero expected profit constraints. This also partially explains why this is an equilibrium (ignoring the integer constraint). The equilibrium contract maximizes active entrepreneurs' expected profits given the price P_G^* and the lenders zero expected profit constraint. So any other contract would result in the deviant entrepreneur making expected losses if the contract is accepted if P_G^* results. But, if a deviant entrepreneur

offers a different contract in period 2 that guarantees a lender non-negative expected profits, then (for sufficiently small β) it is an equilibrium in the subsequent period 3 subgame for lenders to accept this contract and a sufficient number of non-deviant contracts to drive the period 4 equilibrium price down to P_G^* . It is "as if" the would-be deviant entrepreneur takes the price as given and thus realizes that any deviant contract that would be accepted would make losses.

This time around, the proof for part (b) where we respect the integer constraint is relatively easy. The idea, similar to before, is to approximate the equilibrium of part (a) by accepting the integer number of firms just smaller than $N_G^*(\beta)$. The candidate equilibrium is then constructed by respecting the market clearing, entrepreneurs' zero expected profit and lender's zero expected profit constraints but slightly relaxing the efficiency condition of Equation 3.2(iv). Since there are slightly too few active firms, this involves slightly increasing the probability of success of each firm above the efficient level for the candidate equilibrium price. This price will generally be slightly above P_G^* . But if any entrepreneur offers a different contract that would be accepted by a lender, for sufficiently large β , it is an equilibrium in the subsequent subgame to accept enough deviants to drive the price down to or slightly below P_G^* . Thus the deviant will make at best zero profit. Again, intuitively, for small β , these candidate equilibria are very close to that described in part (a).

3.3. Equilibria of the Whole Game.

We now move to consider equilibria in the whole game; that is, we allow deviations in the type of financial institutions formed and in the type of loan contracts. Our aim will be to find conditions under which either the A or the G -subgames or both lie on equilibrium paths.

First we define a price, P_B^* , below which, loosely speaking, no lender will accept a non-monitored, non-tradeable debt contract. Formally, let $[R_B, N_B(\beta), \pi_B, P_B]$ be (uniquely) defined by:

$$\begin{aligned} \text{(i)} \quad P_B^* &= \alpha - \beta q_B^* \pi_B^* N_B^*(\beta) & \text{(ii)} \quad P_B^* - \varphi'(\pi_B^*) &= R_B^* \\ \text{(iii)} \quad R_B^* &= vk(q_B^*)/q_B^* \pi_B^* & \text{(iv)} \quad \varphi''(\pi_B^*) &= vk(q_B^*)/q_B^* (\pi_B^*)^2. \end{aligned}$$

That is, P_B^* is defined by the same set of equations that defined P_A^* except that the marginal cost of capital t has been replaced by v . Given Assumption A.1, for sufficiently small β , P_B^* is well defined and is the lowest price that can emerge on an equilibrium path of a period 1 subgame in which there are active entrepreneurs with non-monitored non-traded loan contracts. The proof mirrors that of Proposition 3.1(a). Moreover, since $v > t$, we know that $P_B^* > P_A^*$.

Define the A^* -path of the game starting in period 0 to be that described in

Proposition 3.1. That is: in period 0, all potential lenders choose to form discount houses (A) and all potential entrepreneurs choose q_A^* ; in period 1, all potential entrepreneurs offer the non-monitored, tradeable loan contract $R_A(\beta)$; (in period 2 there are no movers); in period 3, potential lenders accept exactly $N_A(\beta)$ such contracts; and in period 4, the auctioneer announces the future price $P_A(\beta)$ and all active entrepreneurs choose success probabilities $\pi_A(\beta)$. Define the G^* -path of the game starting in period 0 to be that described in Proposition 3.2. That is: in period 0, all potential lenders choose to form investment banks (G) and all potential entrepreneurs choose q_G^* ; (in period 1, there are no movers); in period 2, all potential entrepreneurs offer the monitored, loan contract $(R_G(\beta), \pi_G(\beta))$; in period 3, lenders accept exactly $N_G(\beta)$ such contracts; and in period 4, the entrepreneur announces the price $P_G(\beta)$ and all active entrepreneurs implement $\pi_G(\beta)$.

The next proposition gives necessary and sufficient conditions for these to be equilibrium paths for small β .

Proposition 3.3. *Given Assumptions A.1 and A.2, for small enough β ,*

(a) *A necessary condition for A^* -path to be an equilibrium path is that*

$$(i) \quad q_A^* \left(\pi P_A^* - \varphi(\pi) \right) - [vk(q_A^*) + q_A^* m + M] \leq 0, \text{ for all } \pi \in [0, 1].$$

The same condition with strict inequality is sufficient. A simpler sufficient condition is that $P_G^ \geq P_A^*$.*

(b) *A sufficient condition for the G^* -path to be an equilibrium path is that*

$$(ii) \quad P_G^* < P_B^*.$$

A necessary condition is that $P_G^ \leq P_B + \beta q_A^* \pi_B$. If we require one necessary condition to hold for all (small) β then this reduces to $P_G^* \leq P_B$.*

Proof: See Appendix.

A loose intuition ignoring the integer constraint is as follows. To sustain the A^* -path as an equilibrium path, whenever an investment bank is established, any deviant potential entrepreneur who offers a monitored loan contract must make losses if accepted. We know from Proposition 3.2 that if the future price is weakly less than P_G^* then at best the entrant will make zero expected profits. But the deviant firm was set up for a Anglo-Saxon type loan. That is, in period 0, not expecting a lender to deviate, the deviant entrepreneur chose scale q_A^* : the "wrong" scale for a monitored loan. Therefore, if the future price falls to P_G^* or below, the entrant will make expected losses.

The trick is to drive the price down this far. Using the idea of Proposition 3.1(a), we know that lenders can respond to the deviation by accepting just enough non-deviant contracts in addition to the deviant to drive the price down to the putative equilibrium

price, P_A^* . But this is the lowest price at which lenders are willing to accept non-deviants. Therefore, if $P_A^* \leq P_G^*$ then the deviant entrepreneur would make losses if she entered. The more complex condition allows for the "slack" arising from deviants having inefficient scale. It is "as if" the would-be deviant entrepreneur took the equilibrium price as given when deciding whether or not to enter with a monitored loan. Figure 3.3(i) illustrates this intuition for the special case where $P_A^* = P_G^*$.

The intuition for the G^* -equilibrium path is a little different. Whenever a non-monitored loan contract is accepted as part of an equilibrium of a period 4 subgame, the lender must know that the entrepreneur expects to make positive profit at the subsequent subgame equilibrium price. Otherwise the entrepreneur would set the probability of success equal to zero and the lender would not get repaid. Any accepted non-monitored entrepreneur then makes profit. Therefore, to sustain the G^* -path as an equilibrium path, any non-monitored loan contract offered by a deviant must be rejected.

If there is only one non-monitored loan contract then markets will be thin so the relevant deviant loan type is both non-monitored and non-traded. By the definition of P_B^* , for such a loan contract to be accepted requires a subsequent subgame equilibrium price $P \geq P_B^*$. Suppose then that an entrepreneur deviates and offers such a loan contract and suppose that lenders rejected the deviant's offer and continued to accept exactly the same number of non-deviant monitored loan contracts as specified by the G -path. In this case, the equilibrium price in the subsequent subgame is approximately P_G^* , the putative equilibrium price. If a deviant lender had accepted the deviant entrepreneur's offer then the subsequent price would have been even lower. Thus, if $P_G^* \leq P_B^*$ then no lender will accept a deviant non-monitored loan contract. It is "as if" the would-be deviant lender took the number of non-deviant incumbents as given when deciding whether or not to accept a would-be entrant's non-monitored loan contract. Figure 3.3(ii) illustrates this intuition again for the special case where $P_A^* = P_G^*$.

3.4. Interpretation of the Results

We have constructed a formal model with both moral hazard and competitive entry in which both the size and number of industrial firms emerge endogenously alongside a financial system. In our model, firms' sizes are relatively larger on the G^* -equilibrium path with monitored loans than on the A^* -equilibrium path with tradeable debt, that is $q_G^* > q_A^*$. If parameters are such that the equilibrium prices, P_G^* and P_A^* , are the same (that is, expected industrial output is the same), then there are more active firms on the A path; $N_A^* > N_G^*$. This corresponds to the simpler model of Section 2, where we took size and number of firms to be exogenous. At the same expected output, there are equilibria with many small firms funded by Anglo-Saxon style loans and with fewer larger firms funded by German style loans.

Proposition 3.3 gives conditions under which our putative A and G equilibria exist. It is perhaps easier to see what this means when we consider what equilibrium paths can be sustained as we vary parameters, ignoring the integer constraint. Consider, for example, changing the fixed cost of monitoring, M , holding other parameters (including small β) fixed. Let $P_G^*(M)$ be the future price, P_G^* , described by Proposition 3.2 as a function of M . Since $P_G^*(M)$ is defined to be the future price that just yields zero expected profit, we can think of this as a cost function. As such, we know that it is increasing (and concave) in M . Now, let \underline{M} be the lowest fixed cost of monitoring such that it is possible to sustain an A -equilibrium.⁵¹ That is, $q_A^*[\pi P_A^* - \varphi(\pi)] - [vk(q_A^*) + q_A^*m + M] \leq 0$, for all $\pi \in [0,1]$, if and only if $M \geq \underline{M}$. Let \bar{M} be the highest fixed cost at which a G -equilibrium can be sustained. That is, $P_G^*(\bar{M}) - \beta q_B \pi_B = P_B^*$. And let M^* be such that $P_G^*(M^*) = P_A^*$.

Figure 3.4, then, illustrates the results. Proposition 3.3 allows us to identify four intervals of the fixed cost, M . By definition, for $M < \underline{M}$, we can sustain a G but not an A -equilibrium. While, for $M > \bar{M}$, we can sustain an A but not a G -equilibrium. From Proposition 3.3(a), we know that $P_G^*(\underline{M}) < P_A^*$, thus $\underline{M} < M^*$ as shown. From Proposition 3.3(b), we know that $P_G^*(\bar{M}) > P_B^* > P_A^*$, thus $\bar{M} > M^*$ as shown. This leaves a non-empty interval of fixed costs, $[\underline{M}, \bar{M}]$, over which both the G and the A equilibria can be sustained.

Thus, the model may help explain how it was possible for different systems of industrial finance to survive in different economies; for example, why the British and German systems did not converge. The model permits a range of parameters for which there is both an equilibrium in which lenders choose to form German style financial institutions and entrepreneurs choose to form German style firms, and also an equilibrium in which lenders choose to form Anglo-Saxon financial institutions and entrepreneurs choose to form Anglo-Saxon style firms. Loosely speaking, it is possible for an economy to get stuck in either system if financiers and entrepreneurs do not coordinate. In our model, the given organisation of firms (that is, their scale) could hinder the development of investment banking, but was not the obstacle to switching the other way. There it was the lack of established thick secondary markets that could hinder change.

How are the equilibrium paths welfare ranked? Can we say that the German or Anglo-Saxon equilibrium is always better when both exist? Notice that active entrepreneurs always weakly prefer an A -equilibrium contract since it yields them positive expected profits. All lenders and non-active entrepreneurs are indifferent. Consumers benefit from lower expected prices, P . Thus, over the non-empty interval of fixed cost, $M \in [M^*, \bar{M}]$, the A^* -equilibrium path Pareto Dominates the G^* -equilibria path. However,

⁵¹ For heuristic purposes assume $\underline{M} > 0$.

over the non-empty interval, $M \in [\underline{M}, M^*)$, the equilibria are not Pareto ranked. At least toward the bottom of this range, it is plausible that the gains to consumers would outweigh the losses to entrepreneurs. Thus, there are probably ranges of the parameter M for which social surplus maximization would favour the German style equilibrium.

Contrast this with the simpler model of Section 2 and to that of Dewatripont and Maskin. In the former, there were parameter ranges where it was possible to get stuck in a German style equilibrium when an Anglo-Saxon system would have Pareto dominated, but not vice versa. In the latter, there are parameter ranges where it is possible to get stuck in an Anglo-Saxon style equilibrium when a German system would Pareto dominate, but not vice versa. In the model presented here, it is possible to get stuck in either direction, though only in one direction is the welfare comparison a simple matter of Pareto dominance.

4. DIRECTIONS FOR FUTURE RESEARCH.

The model suggests some immediate directions for further research. First, while we know that German firms were generally larger than their English equivalents in the late 19th century, is this still true today? How about American firms? Within Germany, is it still the case that universal banks focus their attention on larger firms. If size is not the key example of entrepreneurial choice, could it easily be replaced in our model?

Second, our model completely abstracts from equity finance. Public sale of equity was already an important way to fund larger projects by the turn of this century. Suppose an economy develops secondary markets for tradeable debt, as along the A^* -equilibrium path above, but also develops large scale equity markets. Such an economy may be well equipped to finance small scale and very large scale investment projects but, lacking German style investment banks, it may be less able to finance medium to large sized firms. This was the main finding of the MacMillan committee on the 20th century British financial system.

Third, given that we have identified a possible coordination failure, is it possible to interpret vertical integration between financing and producing firms as an attempt round the problem. For example, Hilferding suggested that JP Morgan and his associates acted in some ways like a German bank coordinating production and finance at turn of the century.⁵² Similarly, Baker (1993) suggests that an advantage held by US mid-20th century conglomerates and holding companies such as Beatrice was that they were able to provide internal capital markets.

Fourth, if there are coordination failures between finance and industry in either England or Germany, will the opening of a common market in financial services result in

⁵² Hilferding (1910), p.399,n.18. See also De Long (1991) and Ramirez (1992a and 1992b).

convergence? It is possible that if, for example, England is stuck in a less efficient equilibrium, then firms from Manchester or Birmingham could break the coordination trap by snubbing the City of London and obtaining loans direct from Frankfurt.

BIBLIOGRAPHY:

- Aoki, M., "The Contingent Governance of Team Production: Analysis of Institutional Complementarity," mimeo, Stanford University, (1993).
- Allen, F and D., Gale, "A Welfare Comparison of Intermediaries and Financial Markets in Germany and the US", *European Economic Review*, (March, 1995a)
- Allen, F and D., Gale, "Financial Markets, Intermediaries and Intertemporal Smoothing" mimeo, University of Pennsylvania (1995b).
- Allen, F and D., Gale, "Universal Banking, Intertemporal Risk Smoothing and European Financial Integration", mimeo (1995c).
- Anderson, B.L., "Money and the Structure of Credit in the Eighteenth Century", *Business History*, XII, (1970), pp. 85-101.
- Ashton, T.S. "The Bill of Exchange and Private Banks in Lancashire, 1790-1830", *Economic History Review*, XV, 1 (1945), pp. 25-35;
- Ashton, T.S., *An Economic History of England: The 18th Century*, (London: Methuen, 1955).
- Baker, G.P., "Beatrice: A Study in the Creation and Destruction of Value", *Journal of Finance*, 47(3) (July, 1992), pp. 1081-119.
- Best M, and J. Humphries, "The City and Industrial Decline", in Elbaum B., and W. Lazonick (Eds.), *The Decline of the British Economy*, (Oxford, Clarendon, 1986), pp 223-239.
- Buchinsky M., and B. Polak, "The Emergence of a National Capital Market in England, 1710-1880", *The Journal of Economic History*, LIII, 1, (March, 1993), pp. 1-24.
- Calomiris, C., "The Costs of Rejecting Universal banking: American Finance in the German Mirror", NBER conference paper, (Nov, 1992).
- Calomiris, C. and D. Raff, "The Evolution of Market Structure, Information and Spreads in American Investment Banking", working paper, Illinois (1993).
- Cantillo, M., "A Theory of Corporate Capital Structure and Investment", working paper, Stanford, 1994.
- Collins M., "English Bank Lending and the Financial Crisis of the 1870s", *Business History*, XXXII, 2, 1990, pp. 198-224.
- Collins, M., and P. Hudson, "Provincial Bank Lending: Yorkshire and Merseyside, 1820-1860", *Bulletin of Economic Research*, XXXI, 2, 1979, pp. 69-80;
- Crafts, N.F.R., *British Economic Growth during the Industrial Revolution*, Oxford, (1985).
- Crouzet, F., "Capital Formation in Great Britain during the Industrial Revolution" in F. Crouzet (Ed.), *Capital Formation in the Industrial Revolution*, (London; Methuen), 1963, pp. 162-222.
- Diamond, D., "Financial Intermediation and Delegated Monitoring", *Review of Economic Studies*, 51, (July, 1984), pp. 393-414.

- Diamond, P., "Aggregate Demand Management in Search Equilibrium", *Journal of Political Economy*, 90, 5, (Oct, 1982), pp 881-94.
- De Long, J.B., "Did J.P. Morgan's Men Add Value: an economist's perspective on financial capitalism", in P. Temin, (Ed.) *Inside the Business Enterprise: Historical Perspectives on the Use of Information* (University of Chicago Press, 1991), pp. 205-251.
- Dewatripont, M., and E. Maskin, "Credit and Efficiency in Centralised and Decentralised Economies", Harvard Institute of Economic Research, Discussion Paper 1512, (Aug, 1990).
- Edwards, J. and K., Fischer, *Banks, Finance and Investment in Germany*, Cambridge, (1994)
- Edwards, J and S. Ogilvie, "Universal Banks and German Industrialisation: A Reappraisal", working paper, Cambridge (Feb, 1995).
- Folhin, C., "The Role of Financial Intermediation in Industrial Development: The Case of the German *Kreditbanken*, 1871-1914", Berkeley seminar paper, (Oct, 1993).
- Gatrell, V.A.C., "Labour, Power, and the Size of Firms in Lancashire Cotton in the second quarter of the nineteenth century", *Economic History Review*, 2nd series, XXX, 1, (Feb, 1977), pp. 95-139.
- Gerschenkron, A., *Economic Backwardness in Historical Perspective: A Book of Essays*, (Cambridge, Mass, 1962).
- Guinnane, T., "Agricultural Credit Cooperatives" [?], paper presented to EHA annual conference, Tucson, (Oct. 1993).
- Hilferding, R., *Finance Capital, A Study of the Latest Phase of Capitalist Development*, (first published Berlin, 1910; this edition translated by M. Watnick and S. Gordon, London: RKP, 1981)
- Holmstrom, B., "Financing of Investment in Europe: A Theoretical Perspective," mimeo, M.I.T., (1993):
- Homer, S., *A History of Interest Rates*, (New Brunswick, 1963).
- Hoppitt, J. "Financial Crises in 18th Century England", *Economic History Review*, 2nd series, XXXIX, 1, (March, 1986), pp. 39-58.
- Ingham, G., *Capitalism Divided? The City and Industry in British Social Development*, (London: MacMillan, 1984).
- Kennedy, W. *Industrial Structure, Capital Markets and the Origins of British Economic Decline*, (Cambridge: CUP, 1987).
- Kennedy, W., "Capital Markets and Industrial Structure in the Victorian Economy", in J. Van Helten and Y. Cassis (Eds.), *Capitalism in a Mature Economy: Financial Institutions, Capital Exports, and British Industry, 1870-1939*, (London: ?, 1990), pp. 23-52.
- Landes, D., *The Unbound Prometheus, Technological Change and Industrial Development*

in Western Europe from 1750 to the Present, (Cambridge: CUP, 1969).

- Masuyama, S., "Role of Japanese Capital Markets: The effect of Cross-Shareholdings on Corporate Accountability", paper to NEDO conference on Capital Markets and Company Success, (Nov. 1991).
- Mayer, C., "Stock Markets, Financial Institutions, and Corporate Finance", paper to NEDO conference on Capital Markets and Company Success, (Nov 1991).
- Mokyr, J., "The Industrial Revolution and the New Economic History", in J. Mokyr (Ed.), *The Economics of the Industrial Revolution*, (New Jersey: Rowman and Allanheld, 1985), pp. 1-51.
- Mirowski, P., "The Rise (and Retreat) of a Market: English Joint Stock Shares in the Eighteenth Century", *Journal of Economic History*, XLI, 3, (Sept, 1981), pp. 559-577.
- Neal, L., "The finance of business during the industrial revolution" in R. Floud and D. McCloskey, *The Economic History of Britain since 1700*. vol 1, ch. 7, Cambridge, (1994).
- Neal, L., "The Response of the London Stock Market to New Issues: Canals, Railways and Empire in the 19th Century", paper presented at ASSA meetings (1995).
- Neuburgher, H., and H. Stokes, "German Banks and German Growth, 1885-1913: an empirical view", *Journal of Economic History*, XXXIV, 3, (Sept, 1974), pp. 710-731.
- Patterson, M., and D. Reiffen, "The Effect of the Bubble Act on the Market for Joint Stock Shares", *Journal of Economic History*, L, 1, (March, 1990), pp 163-171.
- Pollard, S., "Fixed Capital in the Industrial Revolution in Britain", in *The Journal of Economic History*, XXIV, 3, (Sept, 1964), pp.299-314.
- Porter, M., "Capital Disadvantage: America's Failing Capital Investment System", *Harvard Business Review*, 80, 3, (1992), pp. 65-82.
- Ramirez, C. "Did J.P. Morgan's Men Add Liquidity: Corporate Investment, Cash Flow and Financial Structure at the turn of the Twentieth Century", paper presented to EEA conference, (March, 1992a).
- Ramirez, C., "Financial Capitalism in the United States and Germany at the Turn of the Century", paper presented to Harvard Economic History workshop, (Oct, 1992b).
- Riesser, J., *The Great German Banks and their Concentration*, translated for the National Monetary Commission, (Washington, Government Printing Office, 1911, first published 1909) pp. 366-7.
- Scherfke, R., "Finance and Economic Development: The Evolution of the Financial Systems in Germany and the UK," mimeo, Brown University, (1994).
- Tilly, R., "Mergers, External Growth, and Finance in the Development of Large Scale Enterprise in Germany, 1880-1913", *The Journal of Economic History*, XLII, 3, (Sept, 1982), pp. 629-658.

- Tilly R. "German Banking, 1850-1914; Development Assistance for the Strong," *Journal of European Economic History*, 15, 1, (Spring, 1986), pp. 113-152.
- Tilly, R., "An Overview of the Role of Large German Banks to 1914", in Y. Cassis (Ed.) *Finance and Financiers in European History, 1880-1960*, (Cambridge: CUP, 1992)
- Tilly R., and R. Fremling, "German Banks, German Growth and Economic History", *The Journal of Economic History*, XXXVI, 2, (June, 1976), pp. 416-424.
- Weinstein, D. and Y. Yafeh, "On the Costs of Universal Banking: Evidence from Changing Main Bank Relations in Japan," mimeo, Harvard University, (1994).
- Yafeh, Y. "Corporate Ownership, Profitability and Bank-Firm Ties: Evidence from the American Occupation Reforms in Japan," mimeo, Harvard University, (1993).

APPENDIX FOR SECTION 2

Formal Derivation of Condition 2.1., Condition 2.2 and associated Figures.

The Non-Monitored Loan Problem (4). The first order conditions for the Anglo-Saxon maximization problem (4) are as follows:

$$\begin{aligned}
 (\lambda) \quad & q \left[\pi(1+r) + (1-\pi)d \right] \geq c_A L \quad (\text{where } \lambda \geq 0, \text{ (with equality if } \lambda > 0)) \\
 (\gamma_0) \quad & w + L - D - q \geq 0 \quad (\text{where } \gamma_0 \geq 0 \text{ (with equality if } \gamma_0 > 0)) \\
 (\gamma_r) \quad & qP + D - (1+r)q \geq 0 \quad (\text{where } \gamma_r \geq 0 \text{ (with equality if } \gamma_r > 0)) \\
 (\gamma_d) \quad & D - qd \geq 0 \quad (\text{where } \gamma_d \geq 0 \text{ (with equality if } \gamma_d > 0)) \\
 (\mu) \quad & q \left[P - \varphi'(\pi) - [(1+r) - d] \right] = 0 \\
 (\pi) \quad & q \left[\lambda[(1+r) - d] - \mu\varphi''(\pi) \right] \begin{cases} \geq \\ = \\ \leq \end{cases} 0 \text{ as } \pi \begin{cases} = 1 \\ \in (0,1) \\ = 0 \end{cases} \\
 (r) \quad & q \left[(\lambda - 1)\pi - \gamma_r - \mu \right] = 0 \\
 (d) \quad & q \left[(\lambda - 1)(1 - \pi) - \gamma_d + \mu \right] = 0 \\
 (L) \quad & \gamma_0 - \lambda c_A \leq 0 \quad (\text{with equality if } L > 0) \\
 (D) \quad & 1 - \gamma_0 + (\gamma_r + \gamma_d) \leq 0 \quad (\text{with equality if } D > 0)
 \end{aligned}$$

Since $q > w$ and $D \geq 0$, condition (γ_0) implies that $L > 0$. Thus, condition (L) holds with equality. Adding conditions (r) and (d) yields

$$(\lambda - 1) = (\gamma_r + \gamma_d) \geq 0. \quad (A1)$$

Thus $\lambda > 0$ and condition (λ) binds. Since $c_A > 1$, condition (L) implies that $\gamma_0 > \lambda$. This implies both that condition (γ_0) binds, and (from A1) that $\gamma_0 > 1 + (\gamma_r + \gamma_d)$. So, condition (D) holds with strict inequality, implying $D = 0$. That is, no money is placed into the bank. Condition (γ_0) then gives $L = q - w$; that is, there is no over-borrowing. And condition (γ_d) gives $d \leq 0$; there is no positive repayment if the project fails.

Condition (λ) now reads $q[\pi(1+r) + (1-\pi)d] = c_A(q-w)$. Since the right side is positive and $d \leq 0$, this implies that $(1+r) > 0$ and $\pi \neq 0$. Recall that $\varphi'(1) = \infty$. Therefore, Condition (μ) implies that if $\pi = 1$ then $(1+r) < d$: a contradiction. That is $\pi \in (0,1)$. Since $(1+r) > d$, Condition (π) implies that $\mu > 0$. Therefore, condition (d) implies that $\gamma_d > 0$; that is $d = 0$.

Rewriting obtains

$$(1+r) = c_A(q-w)/q\pi; \quad (A2)$$

the hyperbola shown in Figure 2.1(i). And Condition (μ) now reads

$$P - \varphi'(\pi) = (1+r). \quad (A3)$$

Combining (A2) and (A3) gives Condition 2.1(i). Since $\varphi'(1) = \infty$, this implies $\pi < 1$. Condition (π) is thus an equality and, since $\varphi''(\pi) > 0$ for $\pi > 0$, it implies that $\mu > 0$. Condition (d) and (A1) then gives $\gamma_d > 0$, hence $\lambda > 1$. From (A3), we have $(1+r) < P$, hence condition (γ_r) does not bind: $\gamma_r = 0$. Condition (r) then gives

$$\mu = (\lambda - 1)\pi, \quad (A4)$$

and condition (d) gives

$$\gamma_d = (\lambda - 1). \quad (A5)$$

We can now rewrite condition (π) as $(\lambda-1)\pi\varphi''(\pi) = \lambda(1+r)$. Combining this with (A1) and the fact that $(\lambda-1) > 0$ and using subscripts to indicate solutions, we get:

$$P - \varphi'(\pi_A) = (1+r_A) = c_A(q-w)/q\pi_A \quad 2.1(i)$$

$$\varphi''(\pi_A) \geq (1+r_A)/\pi_A = c_A(q-w)/q\pi_A^2. \quad 2.1(ii)$$

(The weak inequality in 2.1(ii) comes from taking λ to ∞ . This occurs when costs are such that there is only one feasible non-monitored contract).

The Monitored loan problem (3). The first order conditions for the German maximization problem (3) are similar except that, since there is no IC constraint, we can ignore constraint (μ) and set $\mu = 0$ in the other first order conditions. That is, we get the same first order conditions as above except that the cost subscript is different and

$$(\pi)' \quad q \left[P - \varphi'(\pi) + (\lambda - 1)((1+r) - d) \right] \begin{cases} \geq \\ = \\ \leq \end{cases} 0 \text{ as } \pi \begin{cases} = 1 \\ \in (0,1) \\ = 0 \end{cases}$$

Again we assume that a solution to the problem exists. The arguments to show that $\lambda \geq 1$, $D = 0$, $L = q-w$, $d \leq 0$, $(1+r) > 0$ and $\pi \neq 0$ are similar to before. It is easy to show that if $\pi = 1$ then the entrepreneur makes negative expected net profits. Hence $\pi \in (0,1)$.

Let π^* be given by $P - \varphi'(\pi^*) = 0$. From Condition (π)', we know that $\pi \geq \pi^*$ and that, if $\pi > \pi^*$ then $(\lambda-1) > 0$. But Conditions (r) and (d) would then imply that $\gamma_r, \gamma_d > 0$; and hence that $(1+r) = P$ and that $d = 0$. Condition (λ) then implies that $\pi P = \pi(1+r) = c_G(q-w)$, but this again implies that the entrepreneur makes expected net losses.

Therefore, $\pi_G = \pi^*$ as in Condition 2.1(iv). Condition (π)' now implies that $(\lambda-1) = 1$; hence that $\gamma_r = \gamma_d = 0$ and $\gamma_0 = 1$. Notice that r and d are not uniquely determined in this problem but satisfy: $\pi_G(1+r_G) + (1-\pi_G)d_G = c_G(q-w)/q$ and $d_G \leq 0 < (1+r_G) \leq P$.

Comparative Statics: c_A and c_G . The argument of Figure 2.1(iii) that raising base rates favours monitored loans at the margin is shown formally by noting that $\lambda_A > 1 = \lambda_G$.

Condition 2.2: The Non-Monitored Loan Problem (6). The first order conditions for the Anglo-Saxon maximization problem (6) are as follows:

$$\begin{aligned}
 (\lambda) \quad & q[\pi P - \varphi(\pi) - \pi(1+r) - (1-\pi)d] + D \geq 0 && \text{(where } \lambda \geq 0, \text{) (with equality if } \lambda > 0 \text{)} \\
 (\gamma_0) \quad & w + L - D - q \geq 0 && \text{(where } \gamma_0 \geq 0 \text{) (with equality if } \gamma_0 > 0 \text{)} \\
 (\gamma_r) \quad & qP + D - (1+r)q \geq 0 && \text{(where } \gamma_r \geq 0 \text{) (with equality if } \gamma_r > 0 \text{)} \\
 (\gamma_d) \quad & D - qd \geq 0 && \text{(where } \gamma_d \geq 0 \text{) (with equality if } \gamma_d > 0 \text{)} \\
 (\mu) \quad & q\left[P - \varphi'(\pi) - [(1+r) - d]\right] = 0 \\
 (\pi) \quad & q\left[[(1+r) - d] - \mu\varphi''(\pi)\right] \begin{cases} \geq \\ = \\ \leq \end{cases} 0 \text{ as } \pi \begin{cases} = 1 \\ \in (0,1) \\ = 0 \end{cases} \\
 (r) \quad & q\left[(1-\lambda)\pi - \gamma_r - \mu\right] = 0 \\
 (d) \quad & q\left[(1-\lambda)(1-\pi) - \gamma_d + \mu\right] = 0 \\
 (L) \quad & \gamma_0 - c_A \leq 0 && \text{(with equality if } L > 0 \text{)} \\
 (D) \quad & \lambda - \gamma_0 + (\gamma_r + \gamma_d) \leq 0 && \text{(with equality if } D > 0 \text{)}
 \end{aligned}$$

We assume a solution exists. Since $q > w$ and $D \geq 0$, Condition (γ_0) implies that $L > 0$. Thus Condition (L) holds with equality. Adding conditions (r) and (d) yields $(1-\lambda) = (\gamma_r + \gamma_d) \geq 0$. Hence $\lambda \leq 1$. Substitution in Condition (D) yields $1-\gamma_0 \leq 0$. Using Condition (L) , $1-c_A \leq 0$ but this inequality is strict by assumption. Hence $D = 0$. Condition (γ_0) then gives $L = q - w$; that is, there is no over-borrowing. And condition (γ_d) gives $d \leq 0$; there is no positive repayment if the project fails. For the lender to make expected profit thus implies $(1+r) > 0$.

If $\pi_A = 0$, then since $d \leq 0$, the lender makes expected losses: a contradiction. If $\pi = 1$, then Condition (μ) implies $(1+r) < 0$ and again the lender would make expected losses. Therefore, $\pi \in (0,1)$. Condition (π) then implies that $\mu > 0$. This, in turn, implies by Condition (d) that $\gamma_d > 0$ and hence that $d = 0$.

Next we show that Condition (λ) does not bind. Suppose (contrahypothesis) that $\pi P - \varphi(\pi) - \pi(1+r) = 0$. By Condition (μ) , $P - \varphi'(\pi) = (1+r)$. Substitution gives $\pi P - \varphi(\pi) - \pi[P - \varphi'(\pi)] = 0$. Rearranging gives $\pi\varphi'(\pi) = \varphi(\pi)$ which contradicts $\pi > 0$, $\varphi(0) = 0$ and $\varphi'' > 0$. Hence $\lambda = 0$.

Since $P - \varphi'(\pi) = (1+r)$, and $\pi > 0$, $P > (1+r)$. Hence, $\gamma_r = 0$. Condition (r) now implies that $\mu = \pi$. So Condition (π) reduces to

$$(1+r_A^L)/\pi_A^L = \varphi''(\pi_A^L).$$

But this is just the tangency illustrated in Figure 2.1(iv).

We claim that $r_A^L \geq r_A$ or, equivalently, $\pi_A^L \leq \pi_A$. We know from above that: $P - \varphi'(\pi_A) = (1+r_A)$ and $\pi_A \varphi''(\pi_A) \geq (1+r_A)$. And we now also know that: $P - \varphi'(\pi_A^L) = (1+r_A^L)$ and $(1+r_A^L) = \pi_A^L \varphi''(\pi_A^L)$. Suppose (contrahypothesis) that $r_A^L < r_A$. This is equivalent to $\pi_A^L > \pi_A$. Then, since $\varphi''' > 0$,

$$(1+r_A^L) = \pi_A^L \varphi''(\pi_A^L) > \pi_A \varphi''(\pi_A) \geq (1+r_A)$$

which is a contradiction.

The Monitored Loan Problem (5). The first order conditions are the same except that Condition (μ) does not apply, $\mu = 0$ in all other conditions; the costs are c_G ; and Condition (π) becomes

$$(\pi)' \quad q[(1-\lambda)((1+r)-d) + \lambda(P-\varphi'(\pi))] \begin{cases} \geq \\ = \\ \leq \end{cases} 0 \text{ as } \pi \begin{cases} = 1 \\ \in (0,1) \\ = 0 \end{cases}$$

Again, we assume a solution exists. The arguments to show that $\lambda \leq 1$, $D = 0$, $L = q-w$, $d \leq 0$, $(1+r) > 0$ and $\pi \neq 0$ are similar to before. If $\pi = 1$ then Condition (λ) is violated. Hence $\pi \in (0,1)$.

Suppose (contrahypothesis) $\lambda = 0$ then, from Conditions (r) and (d), we know that $\gamma_r, \gamma_d > 0$. This would imply $P = (1+r)$ and $d = 0$. But then substitution into Condition (λ) obtains a contradiction.

Let π^* be given by $P - \varphi'(\pi^*) = 0$. From Condition (π)', we know that $\pi \geq \pi^*$ and that, if $\pi > \pi^*$ then $(1-\lambda) > 0$. But Conditions (r) and (d) would then imply that $\gamma_r, \gamma_d > 0$; and hence that $(1+r) = P$ and that $d = 0$. Then Condition (λ) will again yield a contradiction. Hence $\pi_G^L = \pi_G = \pi^*$ (and $\lambda_G = \lambda_G = 1$) as shown in Figure 2.1(iv).

Comparing across bargain powers: Since $\pi_A^L \leq \pi_A$ while $\pi_G^L = \pi_G$, we know that the dead-weight loss is no smaller in Figure 2.1(iv) than in Figure 2.1(i); that is, the dead-weight loss associated with non-monitored loans is weakly increased when we transfer the bargaining power from borrowers to lenders. Since there was (strict) slack in Condition (λ) in the lenders' non-monitored loan problem, we have also shown that not all the surplus from a non-monitored loan was transferred with the bargaining power. Either of these is sufficient to show that transferring the bargaining power to lenders favours monitored loans.

Comparative Statics for Section 2.2. and Associated Figures.

Let $\pi(c,w)$ be the optimal non-monitored loan probability of success when the marginal cost of such a loan is c and the wealth of the entrepreneur is w . Let $z(c_A, w) := [\pi_G^L P - \psi(\pi_G^L)] - [\pi(c,w)P - \psi(\pi(c,w))]$ be the associated dead-weight loss. Recall that we assume all projects are the same size.

Figure 2.2(i): By assumption, there are no other securities traded on the capital market except those issued by firms and borrowers have no private wealth of their own. Let $\hat{q}(0)$ be the project size such that the firm is indifferent between a monitored and a non-traded, non-monitored loan: $M+m\hat{q}(0) = z(v,0)\cdot\hat{q}(0)$. The boundary where non-monitored tradeable loans are just indifferent to non-monitored non-tradeable loans is given by:
 $f(N)+F/Nq+l = v$.

For $q \geq \hat{q}(0)$, let $N(q,0)$ denote the number of firms trading on the capital market such that borrowers of size q are indifferent between taking out a tradeable, non-monitored loan and a monitored bank loan. That is, $N(\cdot,0)$ represents the combinations of N and q where borrowers are just indifferent between monitored and non-monitored, tradeable loans. Therefore, $N(q,0)$ solves $M+mq+νq = q[z(t(N,q))+t(N,q)]$. Differentiating the function $N(q,0)$ and rearranging yields

$$N_q \cdot [f' - F/(Nq^2)] \cdot (1+z_c) = -[M - (1+z_c)F/Nq]/q^2.$$

Notice that at $\hat{q}(0)$, $t(N(\hat{q}(0)),\hat{q}(0)) = v$. Hence the condition for the slope of $N(q,0)$ to be positive at $\hat{q}(0)$ is just:

Condition 2.2.1. $M > ((1+z_c(v,0))F/N(\hat{q}(0),0)\hat{q}(0))$.

As discussed in the text, this assumption holds if fixed costs in monitoring are sufficiently greater than those in capital markets. We now show that this condition is sufficient for the slope to continue to be positive above $\hat{q}(0)$.

We first show that $z_c, z_{cc} > 0$. From the definition of z ,

$$z_c = -[P - \psi'(\pi)]\pi_c \quad \text{and} \\ z_{cc} = -[-\psi''(\pi)(\pi_c)^2 + (P - \psi'(\pi))\pi_{cc}].$$

From the first order conditions for the non-monitored loan problem (with $w=0$) we have $\pi(P - \psi'(\pi)) = c$ which implies

$$\pi_c [P - \psi'(\pi) - \pi\psi''(\pi)] = 1$$

And from the second order condition:

$$P - \psi'(\pi) - \pi\psi''(\pi) < 0,$$

implying that $\pi_c < 0$ and $z_c > 0$.

Differentiating again yields,

$$\pi_{cc} = [2\psi''(\pi) + \pi\psi'''(\pi)]/[P - \psi'(\pi) - \pi\psi''(\pi)] < 0.$$

It follows that $z_{cc} > 0$.

That is, as q rises from \hat{q} , given Condition 2.2.1, initially N rises. Therefore $t(N,q)$ falls and so does z_c . Thus, as we raise q , the right side of Condition 2.2.1 falls and the slope continues to be upward.

Figure 2.3. Including traded assets other than those of the firms, simply reduces the cost of traded loans at every (N, q) . Therefore, the border $N(q, 0)$ and that between traded and non-traded non-monitored loans are lowered everywhere.

Figure 2.4. Now we allow for entrepreneurs to have some private wealth, w . As before, let $\hat{q}(w)$ be the project size such that the firm is indifferent between a monitored and a non-traded, non-monitored loan: $M + m\hat{q}(w) = z(v, w) \cdot \hat{q}(w)$. Since $z_w < 0$ (as long as $w < q$), it follows that $\hat{q}_w(w) > 0$. Similarly, the border between traded and non-traded non-monitored loans is given by $t(N, (q-w)) = f(N) + F/N(q-w) + l = v$, so it is clear that it is shifted upward as we increase w .

The difficult case is the border $N(q, w)$ between traded and monitored loans; that is the solution to: $m \cdot q + v \cdot (q-w) + M = q \cdot z[t(N, (q-w))] + t(N, (q-w)) \cdot (q-w)$ for $q \geq \hat{q}(w)$. Differentiation yields,

$$N_w \cdot [((q-w) + qz_c)t_N] = (t-v) - qz_w - ((q-w) + qz_c)t_w.$$

Since $z_c > 0$ and $t_N < 0$, the sign of N_w is the opposite to the sign of the right side.

For $q > \hat{q}(w)$, and $N = N(q, w)$, we know that traded loans are preferred to non-monitored non-traded loans, hence $(t-v) < 0$. That is, the direct cost of borrowing less as we increase w favours monitored loans. Similarly, $t_w > 0$: increasing wealth makes secondary markets thinner and hence traded loans more costly. But, $z_w < 0$; that is, borrowing less reduces the dead-weight loss of non-monitored loans favouring. The total effect is thus ambiguous.

APPENDIX FOR SECTION 3

Proof of Proposition 3.1:

(a) First, we will show that there exists a unique solution to equations 3(a.i) through 3(a.iv) for sufficiently small β if we do not restrict N to be an integer. Then, we will argue that there can be no lower output price P on an equilibrium path of a period 1 subgame in which there are active firms with non-monitored loan contracts. Finally, we will show that this path is an equilibrium path.

Since $\varphi''(0) = 0$, $\varphi''(1) = \infty$, and $\varphi''' > 0$, there exists a unique π_A^* such that $\varphi''(\pi_A^*) = tk(q_A^*)/q_A^*(\pi_A^*)^2$. Substitution into first 3(a.iii) then 3(a.ii) then 3(a.i) yields unique solutions for R_A^* , P_A^* and $N_A^*(\beta)$ respectively. The expression that defines $N_A^*(\beta)$ is

$$\alpha - \beta q_A^* \pi_A^* N_A^*(\beta) - \varphi'(\pi_A^*) = tk(q_A^*)/q_A^* \pi_A^* \quad (1)$$

So far nothing ensures that the solution $N_A^*(\beta)$ is positive let alone large enough to ensure that $N_A^*(\beta)k(q_A^*) \geq V$; that is, such that the marginal cost of tradeable debt is $t < v$. This is where we need Assumption A.1(a) and our hypothesis that β is small. Notice that π_A^* maximizes the expression $\alpha - \varphi'(\pi) - tk(q_A^*)/q_A^* \pi$. So, since $t < v$,

Assumption A.1(a) implies that $\alpha - \varphi'(\pi_A^*) - tk(q_A^*)/q_A^*\pi_A^* > 0$. Thus, by choosing β small enough in equation (1) we can set $N_A^*(\beta)$ arbitrarily large. In particular we can set $N_A^*(\beta) > N_V := \text{Min}\{N \in \mathbb{Q} \mid N_V k(q_A^*) \geq V\}$.

Next, suppose contrahypothesis that some price $P < P_A^*$ is announced on an equilibrium path of a period 1 subgame such that some non-monitored loan contract, R , is accepted. In this equilibrium, the entrepreneur with this contract must choose probability of success, π , given by $P - \varphi'(\pi) = R$. And, for the contract to have been accepted requires that $R \geq tk(q)/q_A^*\pi$. That is, for this to be an equilibrium, there must exist a π such that $P - \varphi'(\pi) - tk(q_A^*)/q_A^*\pi \geq 0$. But, again notice that π_A^* maximizes the left-side of this expression and, by definition, $P_A^* - \varphi'(\pi_A^*) - tk(q_A^*)/q_A^*\pi_A^* = 0$. Hence $P - \varphi'(\pi) - tk(q_A^*)/q_A^*\pi < 0$ for all π .

Intuitively, a similar argument supports existence of the equilibrium. We claim that, if any one potential entrepreneur deviates in period 1, it is an equilibrium of the subsequent period 3 subgame for potential lenders to accept exactly $N_A^*(\beta)$ non-deviant contracts. We have to show that no potential lender would accept either the deviant contract or accept an extra non-deviant contract, given that $N_A^*(\beta)$ non-deviants are accepted.

To do this and also develop notation for part (b), consider the period 4 subgame following acceptance in period 3 of N firms with contract R and size q_A^* and one "deviant" firm also of size q_A^* . Consider a putative equilibrium of this subgame in which the deviant firm chooses probability of success $\hat{\pi}$. Let $\tilde{P}(\hat{\pi}, R, N, \beta)$ be the price announced by the auctioneer in this putative equilibrium and let $\tilde{\pi}(\hat{\pi}, R, N, \beta)$ be the equilibrium probability of success level that would be chosen by each of the N entrepreneurs with contract R . Formally, $\tilde{\pi}(\hat{\pi}, R, N, \beta) := \max(\{\pi \in [0, 1] \mid \alpha - \beta q_A^* \hat{\pi} - \beta q_A^* N \pi - \varphi'(\pi) = R\} \cup \{0\})$ and $\tilde{P}(\hat{\pi}, R, N, \beta) := \alpha - \beta q_A^* \hat{\pi} - \beta q_A^* N \tilde{\pi}(\hat{\pi}, R, N, \beta)$.

The function \tilde{P} is strictly decreasing in $\hat{\pi}$, decreasing in N and increasing in R . To see this, suppose contrahypothesis that there exist $R, N, \beta, \hat{\pi}^H$, and $\hat{\pi}^L$ such that $\hat{\pi}^H > \hat{\pi}^L$ and $\tilde{P}(\hat{\pi}^H, R, N, \beta) \geq \tilde{P}(\hat{\pi}^L, R, N, \beta)$. Then, by definition, $\tilde{\pi}(\hat{\pi}^H, R, N, \beta) \geq \tilde{\pi}(\hat{\pi}^L, R, N, \beta)$ and hence $\tilde{P}(\hat{\pi}^H, R, N, \beta) < \tilde{P}(\hat{\pi}^L, R, N, \beta)$; a contradiction. Similarly, if contrahypothesis there exist $R, \beta, \hat{\pi}, N^H$ and N^L such that $N^H > N^L$ and $\tilde{P}(\hat{\pi}, R, N^H, \beta) > \tilde{P}(\hat{\pi}, R, N^L, \beta)$ then, by definition, $\tilde{\pi}(\hat{\pi}, R, N^H, \beta) \geq \tilde{\pi}(\hat{\pi}, R, N^L, \beta)$ and hence $\tilde{P}(\hat{\pi}, R, N^H, \beta) \leq \tilde{P}(\hat{\pi}, R, N^L, \beta)$. And, if contrahypothesis there exist $N, \beta, \hat{\pi}, R^H$ and R^L such that $R^H > R^L$ and $\tilde{P}(\hat{\pi}, R^L, N, \beta) > \tilde{P}(\hat{\pi}, R^H, N, \beta)$ then, by definition, $\tilde{\pi}(\hat{\pi}, R^L, N, \beta) \geq \tilde{\pi}(\hat{\pi}, R^H, N, \beta)$ and hence $\tilde{P}(\hat{\pi}, R^L, N, \beta) \leq \tilde{P}(\hat{\pi}, R^H, N, \beta)$.

For the deviant firm to have been accepted by a lender in period 3, it must be the case that its non-monitored loan contract, \hat{R} satisfy $\hat{R} \geq tk(q_A^*)/q_A^*\hat{\pi}$. But $\hat{\pi}$ is given by the extra firms equilibrium first order condition, $\tilde{P}(\hat{\pi}, R, N, \beta) - \varphi'(\hat{\pi}) = \hat{R}$. Thus, a necessary condition for the firm to have been accepted is that there exist a $\hat{\pi}$ such that

$$\tilde{f}(\hat{\pi}, R, N, \beta) := \tilde{P}(\hat{\pi}, R, N, \beta) - \varphi'(\hat{\pi}) - tk(q_A^*)/q_A^* \hat{\pi} \geq 0;$$

or equivalently that there exist a $\hat{\pi}$ such that $\tilde{P}(\hat{\pi}, R, N, \beta) \geq P_A^*$.

Now suppose that, contrahypothesis, an extra firm is accepted by some lender in addition to the $N_A^*(\beta)$ non-deviant firms accepted by other lenders. For this firm to have been accepted requires that there exist a $\hat{\pi}$ such that $\tilde{P}(\hat{\pi}, R_A^*, N_A^*(\beta), \beta) \geq P_A^*$. But, by definition, $\tilde{P}(0, R_A^*, N_A^*(\beta), \beta) = P_A^*$ and \tilde{P} is strictly decreasing in $\hat{\pi}$.

Finally notice that, if lenders accept exactly N_A^* entrepreneurs with the contract R_A^* in period 3 then, by equations 3(ai) and 3(aii) it is an equilibrium of the subsequent period 4 subgame for the auctioneer to announce prices P_A^* and active firms to choose success probabilities π_A^* .

And given equation 3(a.iii), lenders are then just indifferent between accepting and rejecting the equilibrium contracts in period 3. \square

(b) Let $N_A^+(\beta) := \min \{N \in \mathbb{Q} \text{ st } N \geq N_A^*(\beta)\}$ and $N_A^-(\beta) := \max \{N \in \mathbb{Q} \text{ st } N \leq N_A^*(\beta)\}$ be respectively the smallest integer not less than and the largest integer not greater than $N_A^*(\beta)$. We are going to show that, for small enough β , there either exists an equilibrium with $N_A^-(\beta)$ active firms or one with $N_A^+(\beta)$ active firms. First, we construct the corresponding candidate equilibrium paths. Then we prove a series of lemmas concerning the properties of the function \tilde{f} defined in part (a) above. Finally, we use these lemmas to show that at least one of the candidate equilibrium paths are indeed equilibria.

Let $g(\pi, N, \beta) := \alpha - \beta q_A^* N \pi - \varphi'(\pi) - tk(q_A^*)/q_A^* \pi$. Let $\pi(N, \beta)$ solve $g(\pi, N, \beta) = 0$ where $g_{\pi}(\pi, N, \beta) \leq 0$. By Assumption 1(a), given any N , such a solution exists for sufficiently small β . Let $R(N, \beta) := tk(q_A^*)/q_A^* \pi(N, \beta)$ and let $P(N, \beta) := \alpha - \beta q_A^* N \pi(N, \beta)$. Notice that $g_{\pi\pi}(\pi, N, \beta) < 0$; and that, where they are well defined, $\pi_N(N, \beta) < 0$, $R_N(N, \beta) > 0$, and $0 < \pi(N, \beta) < \pi_1$.

Notice in particular that $[R(N_A^*(\beta), \beta), N_A^*(\beta), \pi(N_A^*(\beta), \beta), P(N_A^*(\beta), \beta)] = [R_A^*, N_A^*(\beta), \pi_A^*, P_A^*]$. Loosely speaking, the path $[R(N, \beta), N, \pi(N, \beta), P(N, \beta)]$ defined by setting N to either $N_A^+(\beta)$ or $N_A^-(\beta)$ will be our candidate equilibrium for small enough β . It is clear that both $[R(N_A^+(\beta), \beta), \pi(N_A^+(\beta), \beta), P(N_A^+(\beta), \beta)]$ and $[R(N_A^-(\beta), \beta), \pi(N_A^-(\beta), \beta), P(N_A^-(\beta), \beta)]$ converge to $[R_A^*, \pi_A^*, P_A^*]$ as β goes to zero and that $(N_A^*(\beta)-1) < N_A^-(\beta) \leq N_A^+(\beta) < (N_A^*(\beta)+1)$. It remains to show that, for sufficiently small β , at least one of these paths is always an equilibrium.

It turns out to be useful to define one further function $h(\pi, P) := P - \varphi'(\pi) - tk(q_A^*)/q_A^* \pi$. We know that, for values of N and β such that $\pi(N, \beta)$ is well-defined, $h[\pi(N, \beta), P(N, \beta)] = 0$. From part (a), we also know that P_A^* is the smallest price, P , such that there exists a π such that $h[\pi, P] = 0$; that is $h_{\pi}[\pi_A^*, P_A^*] = 0$. Thus, $P(N, \beta) \geq P_A^*$ for all N and β . Since $\pi_N(N, \beta) < 0$, and $h_{\pi\pi}(\pi, P) < 0$, it follows that $h[\pi, P(N_A^+(\beta), \beta)] < 0$ for all $\pi < \pi(N_A^+(\beta), \beta)$. Notice also that, for all $N, \beta > 0$ such that $\pi(N, \beta)$ is defined,

$$h_{\pi}[\pi(N,\beta),P(N,\beta)] > g_{\pi}[\pi(N,\beta),N,\beta].$$

The following lemma defines $\bar{N}(\beta)$.

Lemma 3.1.1. *Given Assumption A.1(a), for each $\beta \geq 0$ sufficiently small there exists a unique $\bar{N}(\beta) > N_A^*(\beta)$ such that $g_{\pi}[\pi(\bar{N}(\beta),\beta),\bar{N}(\beta),\beta] = 0$.*

Proof: Let $V^{\varepsilon}(N,\beta) := \text{Max}_{\pi} g(\pi,N,\beta)$. Then, it is enough to show that (given β , sufficiently small) there is a unique $\bar{N}(\beta)$ such that $V^{\varepsilon}(\bar{N}(\beta),\beta) = 0$. We know that for all $\beta > 0$, $0 = h_{\pi}[\pi_A^*,P_A^*] > g_{\pi}(\pi_A^*,N_A^*(\beta),\beta)$ and $g(\pi_A^*,N_A^*(\beta),\beta) = 0$, hence $V^{\varepsilon}(N_A^*(\beta),\beta) > 0$. By the envelope theorem, $V_N^{\varepsilon}(N,\beta) < 0$ for all $N,\beta > 0$. And, for all $\beta > 0$ there exists an N large enough such that $V^{\varepsilon}(N,\beta) < 0$. \square

Next, consider again a putative equilibrium of a period 4 subgame with $N+1$ active firms all of size q_A^* comprising N "non-deviant" firms with contract R and one "deviant" whose chosen probability of success in this putative subgame equilibrium is $\hat{\pi}$. Recall from part (a) that $\tilde{f}(\hat{\pi},R,N,\beta) := \tilde{P}(\hat{\pi},R,N,\beta) - \varphi'(\hat{\pi}) - tk(q_A^*)/q_A^*\hat{\pi}$. Notice that the function \tilde{f} resembles the function $h(\pi,P)$ except that the price, \tilde{P} , is allowed to change. It also resembles the function $g(\pi,N,\beta)$ except that the price changes due to the change in $\hat{\pi}$ are compensated by changes in $\tilde{\pi}$.

Loosely speaking, the next two lemmas consider the function \tilde{f} where there are $(N-1)$ non-deviants and 1 deviant. Notice in particular that

$$\tilde{f}_{\pi}[\pi(N,\beta),R(N,\beta),(N-1),\beta] = g[\pi(N,\beta),N,\beta] = h[\pi(N,\beta),P(N,\beta)] = 0.$$

Lemma 3.1.2: *Given Assumption 1(a), for each sufficiently small $\beta > 0$ there exists a unique $\tilde{N}(\beta) \in [N_A^*(\beta),\bar{N}(\beta)]$ such that $\tilde{f}_{\pi}[\pi(N,\beta),R(N,\beta),(N-1),\beta] < 0$ for $N \in (N_A^*(\beta),\tilde{N}(\beta))$; $\tilde{f}_{\pi}[\pi(N,\beta),R(N,\beta),(N-1),\beta] > 0$ for $N \in (\tilde{N}(\beta),\bar{N}(\beta))$; and $\tilde{f}_{\pi}[\pi(\tilde{N}(\beta),\beta),R(\tilde{N}(\beta),\beta),(\tilde{N}(\beta)-1),\beta] = 0$.*

Proof: Notice that $\tilde{\pi}[\pi(N,\beta),R(N,\beta),(N-1),\beta] = \pi(N,\beta)$. Simple calculations reveal that:

$$\tilde{\pi}_{\pi}(\hat{\pi},R,(N-1),\beta) = \frac{-\beta q_A^*}{[(N-1)\beta q_A^* + \varphi''(\tilde{\pi}(\hat{\pi},R,(N-1),\beta))]},$$

$$\tilde{P}_{\pi}(\hat{\pi},R,(N-1),\beta) = \varphi''(\tilde{\pi}(\hat{\pi},R,(N-1),\beta)) \cdot \tilde{\pi}_{\pi}(\hat{\pi},R,(N-1),\beta),$$

and
$$\tilde{f}_{\pi}(\hat{\pi},R,(N-1),\beta) = \tilde{P}_{\pi}(\hat{\pi},R,(N-1),\beta) - \varphi''(\hat{\pi}) + tk(q_A^*)/q_A^*\hat{\pi}^2.$$

Thus, for $N > 1$,

$$0 > \tilde{P}_{\pi}[\pi(N,\beta),R(N,\beta),(N-1),\beta] = \frac{-\beta q_A^* \varphi''(\pi(N,\beta))}{[(N-1)\beta q_A^* + \varphi''(\pi(N,\beta))]} > -N\beta q_A^*.$$

Hence,

$$h_{\pi}[\pi(N,\beta),P(N,\beta)] > \tilde{f}_{\pi}[\pi(N,\beta),R(N,\beta),(N-1),\beta] > g_{\pi}[\pi(N,\beta),N,\beta].$$

But, by definition, the left-side is zero at $N_A^*(\beta)$ and the right-side is zero at $\bar{N}(\beta)$.

Therefore, it is enough to show that the middle term is increasing in N . But,

$$\begin{aligned} d[\tilde{f}_{\pi}[\pi(N,\beta),R(N,\beta),(N-1),\beta]]/dN = \\ \left(\frac{-(\beta q_A^*)^2 (N-1) \varphi''(\pi(N,\beta))}{[(N-1)\beta q_A^* + \varphi''(\pi(N,\beta))]^2} - \varphi''(\pi(N,\beta)) - 2tk(q_A^*)/q_A^*(\pi(N,\beta))^3 \right) \cdot \pi_N(N,\beta) + \\ \frac{(\beta q_A^*)^2 \varphi''(\pi(N,\beta))}{[(N-1)\beta q_A^* + \varphi''(\pi(N,\beta))]^2} > 0 \end{aligned}$$

since $\pi_N(N,\beta) < 0$. \square

Lemma 3.1.3: *Given Assumption A.1(b), for sufficiently small β , the function $\tilde{f}[\hat{\pi},R(N,\beta),(N-1),\beta]$ is concave in $\hat{\pi}$ for $N = N_A^+(\beta)$ and for $N = \tilde{N}(\beta)$.*

Proof: Calculation yields: $\tilde{f}_{\pi\pi}[\hat{\pi},R(N,\beta),(N-1),\beta] =$

$$\frac{(\beta q_A^*)^3 (N-1) \varphi''[\tilde{\pi}(\hat{\pi},R(N,\beta),(N-1),\beta)]}{[(N-1)\beta q_A^* + \varphi''[\tilde{\pi}(\hat{\pi},R(N,\beta),(N-1),\beta)]]^3} - \varphi''(\hat{\pi}) - 2tk(q_A^*)/q_A^* \hat{\pi}^3.$$

The last two terms are strictly negative so it is sufficient to show that the first term becomes small as β goes to zero. For sufficiently small β , both $N_A^+(\beta)$ and $\tilde{N}(\beta)$ are greater than one. And, for any $N > 1$ and any $\hat{\pi} \in [0,1]$, $\tilde{\pi}(\hat{\pi},R(N,\beta),(N-1),\beta) \in [0,\pi_1]$. Hence the φ'' term in the numerator is bounded by Assumption A.1(b). It only remains to show that the $\beta N_A^+(\beta)$ or $\beta \tilde{N}(\beta)$ terms in the denominator do not go to zero as β goes to zero. But, notice that $\beta N_A^*(\beta) = [\alpha - P_A^*]/\pi_A^*$ which does not depend on β . And, for all $\beta > 0$, $N_A^+(\beta) \geq N_A^*(\beta)$ by definition and $\tilde{N}(\beta) \geq N_A^*(\beta)$ by Lemma 3.1.2. \square

Using the Lemmas, we can divide our problem into two cases.

Case 1, $\tilde{N}(\beta) > N_A^+(\beta)$: *For sufficiently small β , if $\tilde{N}(\beta) > N_A^+(\beta)$ the path $[R(N_A^+(\beta),\beta),N_A^+(\beta),\pi(N_A^+(\beta),\beta),P(N_A^+(\beta),\beta)]$ forms an equilibrium of the A-subgame.*

Proof: First notice that for sufficiently small β , $1 < N_A^+(\beta) < \bar{N}(\beta)$ and $N_A^+(\beta)q_A^* > V$. Given the proof of part (a) above, it is enough to show that $\tilde{f}[\hat{\pi},R(N_A^+(\beta),\beta),N_A^+(\beta),\beta] \leq 0$ for all $\hat{\pi}$, since this is enough to ensure that it is equilibrium in Period 3 for no contract to be accepted in addition to the $N_A^+(\beta)$ "non-deviants".

First consider $\hat{\pi} \leq \pi(N_A^+(\beta),\beta)$. Recall that $h[\hat{\pi},P(N_A^+(\beta),\beta)] \leq 0$ for all such $\hat{\pi}$. Since \tilde{P} is strictly decreasing in $\hat{\pi}$, we know that for all N and β such that $\pi(N,\beta)$ is defined, $\tilde{P}[\hat{\pi},R(N,\beta),N,\beta] < P(N,\beta)$ for all $\hat{\pi} > 0$; or equivalently, $\tilde{f}[\hat{\pi},R(N,\beta),N,\beta] < h[\hat{\pi},P(N,\beta)]$ for all $\hat{\pi} > 0$. That is, loosely speaking, the period 4 subgame equilibrium price induced by a deviation is less than the non-deviation price. Hence

$\tilde{f}[\hat{\pi}, R(N_A^+(\beta), \beta), N_A^+(\beta), \beta] < 0$ for all $\hat{\pi} \leq \pi(N_A^+(\beta), \beta)$.

Next consider $\hat{\pi} \geq \pi(N_A^+(\beta), \beta)$. Since $\tilde{N}(\beta) > N_A^+(\beta)$, we know from Lemma 3.1.2 that $\tilde{f}_{\pi}[\pi(N_A^+(\beta), \beta), R(N_A^+(\beta), \beta), (N_A^+(\beta)-1), \beta] < 0$ and, from Lemma 3.1.3 that $\tilde{f}_{\pi}[\hat{\pi}, R(N_A^+(\beta), \beta), (N_A^+(\beta)-1), \beta] < 0$ for all $\hat{\pi} \geq \pi(N_A^+(\beta), \beta)$. But, $\tilde{f}[\pi(N_A^+(\beta), \beta), R(N_A^+(\beta), \beta), (N_A^+(\beta)-1), \beta] = 0$. And since $\tilde{f}[\hat{\pi}, R, N, \beta]$ is decreasing in N , $\tilde{f}[\hat{\pi}, R(N_A^+(\beta), \beta), N_A^+(\beta), \beta] \leq \tilde{f}[\hat{\pi}, R(N_A^+(\beta), \beta), (N_A^+(\beta)-1), \beta] \leq 0$ for all $\hat{\pi} \geq \pi(N_A^+(\beta), \beta)$. \square

Case 2, $\tilde{N}(\beta) \leq N_A^+(\beta)$: For sufficiently small β , if $\tilde{N}(\beta) \leq N_A^+(\beta)$ the path $[R(N_A^-(\beta), \beta), N_A^-(\beta), \pi(N_A^-(\beta), \beta), P(N_A^-(\beta), \beta)]$ forms an equilibrium of the A-subgame.

Proof: First notice that for sufficiently small β , $N_A^-(\beta) > 1$ and $N_A^-(\beta)q_A^* > V$. Once again, it is enough to show that $\tilde{f}[\hat{\pi}, R(N_A^-(\beta), \beta), N_A^-(\beta), \beta] \leq 0$ for all $\hat{\pi}$. Recall that $\tilde{P}_N(\hat{\pi}, R, N, \beta) \leq 0$, $\tilde{P}_R(\hat{\pi}, R, N, \beta) \geq 0$, and, where it is defined, $R_N(N, \beta) > 0$. Hence, for all N and N' such that $R(N, \beta)$ and $R(N', \beta)$ are defined and such that $(N+1) \geq N' \geq N$, we know that $\tilde{f}(\hat{\pi}, R(N', \beta), (N'-1), \beta) \geq \tilde{f}(\hat{\pi}, R(N, \beta), N, \beta)$. But, by Lemma 3.1.2, $\tilde{N}(\beta) \geq N_A^*(\beta)$. So, if $\tilde{N}(\beta) \leq N_A^+(\beta)$ then $(N_A^+(\beta)+1) \geq \tilde{N}(\beta) \geq N_A^-(\beta)$. Hence $\tilde{f}[\hat{\pi}, R(N_A^-(\beta), \beta), N_A^-(\beta), \beta] \leq \tilde{f}[\hat{\pi}, R(\tilde{N}(\beta), \beta), (\tilde{N}(\beta)-1), \beta]$.

But, given that $\tilde{f}[\pi(N, \beta), R(N, \beta), (N-1), \beta] = 0$ for all N and β such that $\pi(N, \beta)$ is defined, an immediate consequence of Lemmas 3.1.2 and 3.1.3 is that for sufficiently small β , $\tilde{f}[\hat{\pi}, R(\tilde{N}(\beta), \beta), (\tilde{N}(\beta)-1), \beta] \leq 0$ for all $\hat{\pi}$. Hence we are done. \blacksquare

Proof of Proposition 3.2:

(a) First we argue that, for small enough β , there exists a unique solution to equalities 3.2(a.i) through (a.iv). Let $g^G(\pi, N, \beta) := \alpha - \beta q_G^* N \pi - \varphi'(\pi)$. Since $g_{\pi}^G(\pi, N, \beta) < 0$, let $\pi^G(N, \beta)$ be the unique solution to $g^G(\pi, N, \beta) := 0$. Let $R^G(N, \beta) := [vk(q_G^*) + mq_G^* + M]/q_G^* \pi^G(N, \beta)$ and let $P^G(N, \beta) := \alpha - \beta q_G^* N \pi^G(N, \beta) = \varphi'(\pi^G(N, \beta))$. Notice that $\pi_N^G(N, \beta) < 0$ and hence $R_N^G(N, \beta) > 0$ and $P_N^G(N, \beta) < 0$ for all $\beta > 0$. Moreover, for all $\beta > 0$, $\lim_{N \rightarrow \infty} P^G(N, \beta) = 0$.

Let $h^G(\pi, P) = \pi P - \varphi(\pi)$ be the expected revenue of an active entrepreneur given prices P and with probability of success π , net of private entrepreneurial cost ($\varphi(\pi)$) but gross of expected loan repayments. Let $H^G(P) := \max_{\pi \in [0, 1]} h^G(\pi, P)$. By the envelope theorem, $H_P^G(P) > 0$, hence (abusing notation slightly) $H_N^G(P^G(N, \beta)) < 0$. Moreover, for all $\beta > 0$, $\lim_{N \rightarrow \infty} H^G(P^G(N, \beta)) = 0$. Assumption A.2. implies that $H^G(\alpha) > [vk(q_G^*) + mq_G^* + M]/q_G^*$. Hence, for small enough β (since $\pi \in [0, 1]$), $H^G[P^G(1, \beta)] = H^G[\alpha - \beta q_G^* \pi^G(1, \beta)] > [vk(q_G^*) + mq_G^* + M]/q_G^*$. Therefore, for small enough β (using the monotonicity of $H^G[P(N, \beta)]$ in N), there exists a unique $N_G^*(\beta)$ defined by $H^G[P(N_G^*(\beta), \beta)] = [vk(q_G^*) + mq_G^* + M]/q_G^*$.

Having defined $N_G(\beta)$, we can now define $R_G^* := R^G(N_G^*(\beta), \beta)$; $\pi_G^* := \pi^G(N_G^*(\beta), \beta)$; and $P_G^* := P^G(N_G^*(\beta), \beta)$. Equalities 3.2(a.i) and (a.iii) are then satisfied by construction.

Since $P_G^G(N, \beta) - \varphi'[\pi_G^G(N, \beta)] \equiv 0$, first, P_G^* and π_G^* satisfy equality 3.2(a.iv); and second, $h_G^G(\pi_G^*, P_G^*) \equiv H_G^G(P_G^*) \equiv [vk(q_G^*) + mq_G^* + M]/q_G^*$. But, $h_G^G(\pi_G^*, P_G^*) \equiv \pi_G^* P_G^* - \varphi(\pi_G^*)$ and $[vk(q_G^*) + mq_G^* + M]/q_G^* \equiv R_G^* \pi_G^*$ so equality 3.2(a.ii) is also satisfied.

We next argue that P_G^* is the lowest price that can emerge on an equilibrium path of a period 2 subgame in which there are active monitored entrepreneurs. For any monitored loan contract (π, R) to be accepted by a lender it must satisfy $R\pi \geq [vk(q_G^*) + mq_G^* + M]/q_G^*$. Thus, for the monitored entrepreneur to make non-negative expected profit it must be that the equilibrium price P in the subsequent period 4 subgame is such that $H_G^G(P) \geq [vk(q_G^*) + mq_G^* + M]/q_G^*$. But, by construction, $H_G^G(P_G^*) \equiv [vk(q_G^*) + mq_G^* + M]/q_G^*$ and $H_G^G(P) > 0$. That is, if the equilibrium price were less than P_G^* than any active monitored entrepreneur would make losses: it would not be a period 2 equilibrium to offer this contract knowing that these prices would emerge.

Next, we argue that this is an equilibrium path of the G -subgame ignoring the integer restriction. Equality 3.2(a.iii) ensures that lenders are exactly indifferent and equality 3.2(a.i) ensures that this is the expected market clearing price. Thus lenders accepting $N_G^*(\beta)$ of the proposed contracts and the auctioneer announcing price P_G^* form an equilibrium of the period 3 subgame on the putative equilibrium path. Now suppose that an entrepreneur deviates in period 2 and offers some monitored loan contract $(\hat{\pi}, \hat{R})$. If $\hat{R}\hat{\pi} \leq [vk(q_G^*) + mq_G^* + M]/q_G^*$, then the contract does not offer a lender positive profit so it is an equilibrium in the subsequent period 3 subgame for this contract to be rejected.

Suppose then that $\hat{R}\hat{\pi} > [vk(q_G^*) + mq_G^* + M]/q_G^*$. It is enough to show that there is an equilibrium of the subsequent period 3 subgame in which the deviant contract is accepted and the deviant entrepreneur makes expected losses. Recall that the lenders are indifferent about accepting non-deviant contracts and observe that the subsequent equilibrium price is decreasing in the number of non-deviants accepted. Therefore it is an equilibrium in the period 3 subgame following the deviation for lenders to accept just enough non-deviants along with the deviant such that the subsequent period 4 equilibrium price is driven to exactly P_G^* .⁵⁰ But then we know that the deviant will make losses. \square

(b) We used the lack of an integer restriction twice in the above argument: first to define $N_G^*(\beta)$ and second to construct an equilibrium in the subgame in which the price was exactly P_G^* . Loosely speaking, the definition of $N_G^*(\beta)$ came from the entrepreneurs' zero expected profit condition; equality 3.2(a.ii). Let $N_G^-(\beta)$ be the largest integer weakly less than $N_G^*(\beta)$. Suppose we simply replace $N_G^*(\beta)$ by $N_G^-(\beta)$ in the above definitions. That

⁵⁰ Strictly speaking, it might be that just accepting the deviant already drives the price below P_G^* . In this case an equivalent argument applies.

is, consider the path $P^G(N_G^-(\beta), \beta) \geq P_G^*$, $\pi^G(N_G^-(\beta), \beta) \geq \pi_G^*$, and $R^G(N_G^-(\beta), \beta) \leq R_G^*$. On this path, entrepreneurs may make positive expected profits while lenders make zero expected profits since $\pi^G(N_G^-(\beta), \beta) \cdot P^G(N_G^-(\beta), \beta) - \varphi[\pi^G(N_G^-(\beta), \beta)] \geq R^G(N_G^-(\beta), \beta) \cdot \pi^G(N_G^-(\beta), \beta) \equiv [vk(q_G^*) + mq_G^* + M]/q_G^*$. We will construct a path involving using $N_G^-(\beta)$ active firms in which both entrepreneurs and lenders make zero profit. Loosely speaking, we do this by increasing the probability of success and decreasing the price level until profits are zero.

Let $P^-(\pi) := \alpha - \beta q_G^* N_G^-(\beta) \pi$ and consider $h^G(\pi, P^-(\pi))$. We just argued that $h^G[\pi^G(N_G^-(\beta), \beta), P^-(\pi^G(N_G^-(\beta), \beta))] \geq [vk(q_G^*) + mq_G^* + M]/q_G^*$. For $\pi > \pi^G(N_G^-(\beta), \beta)$, the total differential of $h^G[\pi, P^-(\pi)]$ with respect to π is negative. Hence there exists a $\pi_G(\beta) \geq \pi^G(N_G^-(\beta), \beta)$ such that $h^G[\pi_G(\beta), P^-(\pi_G(\beta))] = [vk(q_G^*) + mq_G^* + M]/q_G^*$. Let $P_G(\beta) := P^-(\pi_G(\beta))$. Notice that $P^G(N_G^-(\beta), \beta) \geq P_G(\beta) \geq P_G^*$. Let $R_G(\beta) = [vk(q_G^*) + mq_G^* + M]/q_G^* \pi_G(\beta)$. Notice that $R^G(N_G^-(\beta), \beta) \leq R_G(\beta) \leq P_G^*$. We are going to use $[R_G(\beta), N_G^-(\beta), \pi_G(\beta), P_G(\beta)]$ as our candidate equilibrium path for small β . By construction, this path satisfies expected market clearing, zero expected entrepreneurs' expected profit and zero lenders' expected profit. That is, it satisfies equalities 3.2(a.i) through (a.iii) but not (a.iv).

It is clear that this path converges to the path $[R_G^*, N_G^*(\beta), \pi_G^*, P_G^*]$ as β goes to zero. It remains to show that this is an equilibrium path. The argument is identical to part (a)

except that following a deviant contract offer by an entrepreneur in period 2, it is not possible to accept exactly enough non-deviants such that the subsequent subgame equilibrium price is P_G^* . Instead, construct an equilibrium in the period 3 subgame following the deviation such that the smallest integer weakly larger than the exact desired number of non-deviants are accepted.

Now the price in the subsequent subgame, call it P' , will typically be slightly below P_G^* and both the deviant and the non-deviant entrepreneurs will make expected losses. But, for small enough β , this is an equilibrium of the period 3 subgame. The entrepreneurs can not defect since by period 3 they are committed to their contracts. The lenders are just indifferent to accept or reject each contract provided the subsequent equilibrium price, P' is above $R_G(\beta)$. But since $P_G^* > R_G^* \geq R_G(\beta)$, for small enough β , $P' > R_G(\beta)$. Thus, for small enough β , there is an equilibrium of the period 3 subgame following the entrepreneur's deviation in which the deviant makes losses. ■

Proof of Proposition 3.3:

(a) Recall that we do not have to deal with simultaneous deviations in any subgame. We shall first deal with deviations in scale. Then we shall argue that condition 3.3(i) with strict inequality is sufficient; then, that $P_A^* < P_G^*$ is sufficient. Finally we shall argue that condition 3.3(i) with weak inequality is necessary.

First, then, consider a potential entrepreneur deviating in period 0 by choosing scale $\hat{q} \neq q_A^*$. We claim that there is an equilibrium in the subsequent period 1 subgame in which, regardless of what the deviant does, all other entrepreneurs offer contract $R_A(\beta)$ and, in the subsequent period 3, the deviant's contract is rejected. $R_A(\beta)$ is as defined in the proof of Proposition 3.1. The proof here is identical except that the argument is strengthened since the deviant here has inefficient scale for a non-monitored loan. A deviation to a more expensive non-monitored non-tradeable loan is ruled out by a similar argument. Given Proposition 3.1, we have now dealt with all deviations that do not involve monitored loans.

Next we show that that strict inequality 3.3(i) is a sufficient condition. Suppose that a potential lender has deviated to G ; that is, formed an investment bank in period 0. We claim that there is an equilibrium of the subsequent period 1 subgame in which all potential entrepreneurs offer the non-monitored loan contract $R_A(\beta)$ as described in the proof of Proposition 3.1, ignoring the deviant investment bank. We know from that proof that this is an equilibrium from period 2 onward. So suppose, contrahypothesis, that an potential entrepreneur deviates by passing (\emptyset) in period 1 and offering a monitored loan contract $(\hat{R}, \hat{\pi})$ in period 2. All potential entrepreneurs committed to scale q_A^* in period 0. Thus, if the deviant contract has $\hat{\pi}\hat{R} \leq [vk(q_A^*) + q_A^*m + M]$, we can construct an equilibrium of the subsequent subgame in which the deviant's contract is rejected by the deviant investment bank.

Suppose, then, that the deviant sets a contract such that $\hat{\pi}\hat{R} > [vk(q_A^*) + q_A^*m + M]$, and consider the subsequent period 3 subgame. In this case, any equilibrium of the subgame must involve the deviant investment bank accepting the deviant contract. We therefore construct an equilibrium of this period 3 subgame in which the deviant entrepreneur is accepted but makes negative expected profit. The idea is to accept just enough non-deviants along with the deviant to drive down the subsequent period 4 equilibrium price approximately to P_A^* .

Recall from the proof of proposition 3.1 that $N_A(\beta)$ solves $g(\pi_A(\beta), N, \beta) = \alpha - \beta q_A^* \pi_A(\beta) N - \varphi'(\pi_A(\beta)) - tk(q_A^*)/q_A^* \pi_A(\beta) = 0$. Given $\hat{\pi}$, let $N'(\hat{\pi}, \beta)$ solve

$$(\alpha - \beta q_A^* \hat{\pi}) - \beta q_A^* \pi_A(\beta) N - \varphi'(\pi_A(\beta)) - tk(q_A^*)/q_A^* \pi_A(\beta) = 0. \quad (1')$$

We claim that $N'(\hat{\pi}, \beta)$ is large when β is small. Assumption A1 implies that, for small enough β , there exists a probability of success $\pi \in [0, 1]$ such that $(\alpha - \beta q_A^* \hat{\pi}) - \varphi'(\pi) - tk(q_A^*)/q_A^* \pi > 0$. But π_A^* maximizes the expression $(\alpha - \beta q_A^* \hat{\pi}) - \varphi'(\pi) - tk(q_A^*)/q_A^* \pi$, so for small enough β , $(\alpha - \beta q_A^* \hat{\pi}) - \varphi'(\pi_A(\beta)) - tk(q_A^*)/q_A^* \pi_A(\beta) > 0$ (since $\pi_A(\beta)$ is close to π_A^* for small β). Thus, by choosing β small enough in equation (1') we can set $N'(\hat{\pi}, \beta)$ arbitrarily large. Notice that this argument shows that, for small β , $(\alpha - \beta q_A^* \hat{\pi}) > P_A^*$ for all $\hat{\pi} \in [0, 1]$.

Recall that the non-deviant entrepreneurs' contracts specify a repayment rate $R_A(\beta) = tk(q_A^*)/q_A^*\pi_A(\beta)$. So, by construction, if lenders were able to accept exactly $N'(\hat{\pi},\beta)$ non-deviant contracts along with the deviant then the subsequent equilibrium price would be exactly P_A^* . But $N'(\hat{\pi},\beta)$ is not an integer. So, instead, we propose a candidate period 3 equilibrium in which lenders accept $N^-(\hat{\pi},\beta)$, defined as the largest integer weakly less than $N'(\hat{\pi},\beta)$. The subsequent equilibrium price, $P^-(\hat{\pi},\beta)$ is weakly greater than P_A^* , and converges to P_A^* as β goes to zero. If one more non-deviant firm were accepted, the subsequent equilibrium price would be less than P_A^* and lenders would make expected losses. Therefore, lenders are indeed willing to accept exactly $N^-(\hat{\pi},\beta)$ in a period 3 equilibrium.

The strict condition 3.3(i) implies that if a deviant firm with a monitored loan has committed to make expected repayments greater than $[vk(q_A^*)+q_A^*m+M]$ and if the period 4 price is P_A^* then the deviant firm makes strictly negative expected profit. So, for small enough β , the price that emerges from our proposed period 3 equilibrium, $P^-(\hat{\pi},\beta)$ is small enough to ensure that the deviant makes expected losses.

The simpler condition $P_G^* \geq P_A^*$ is also sufficient. By the definition of P_G^* in Proposition 3.2, we know that at the optimum scale for a monitored loan contract, q_G^* , $\text{Max}_{\pi} q_G^*[\pi P_G^* - \varphi(\pi)] - [vk(q_G^*)-q_G^*m-M] = 0$. Thus, since q_A^* is not the optimal scale for a monitored loan when $M > 0$, if $P_G^* \geq P_A^*$, then condition 3.3(i) holds.

Finally, we show that weak inequality 3.3(i) is a necessary condition. Suppose that a potential lender deviates and forms an investment bank (G) in period 0. Would a potential entrepreneur (perhaps one whose contract offer is rejected on the putative equilibrium path) make positive profit by passing in period 1 and offering an acceptable monitored loan contract to the investment bank in period 2?

We have argued that for small β , $\alpha - \beta q_A^*\pi > P_A^*$ for all $\pi \in [0,1]$. So if the deviant was the only active entrepreneur in period 4 then prices would be greater than P_A^* . From Proposition 3.1s(a), we know that the lowest price that can result from the non-deviant lenders accepting non-monitored loan contracts in any equilibrium of a period 3 subgame is P_A^* . Thus the lowest price that can emerge from any equilibrium of the period 3 subgame following the deviant monitored loan offer is P_A^* . If the integer restriction is imposed then only a slightly higher price can be achieved. Thus, if there exists a probability of success $\hat{\pi}$, such that, $q_A^*[\hat{\pi}P_A^* - \varphi(\hat{\pi})] - [vk(q_A^*)-q_A^*m-M] > 0$ then a deviant entrepreneur can make positive expected profit while still ensuring the investment bank positive expected returns. \square

(b) Again, we do not have to consider simultaneous deviations. We shall first deal with scale deviations. Then, we shall prove the sufficiency of condition 3.3(ii). Finally, we shall justify the necessary condition.

Suppose that a potential entrepreneur deviates in period 0 by setting scale $\hat{q} \neq q_G^*$ and suppose that all the potential entrepreneurs including the deviant pass in period 1. Consider the subsequent subgame. We claim that there is an equilibrium in the subsequent period 2 subgame in which the non-deviants offer the contract $(R_G(\beta), \pi_G(\beta))$ and the deviant passes again. The contract $(R_G(\beta), \pi_G(\beta))$ is as defined in the proof of Proposition 3.2. The proof is identical except that the argument is strengthened since the deviant here has inefficient scale for a monitored loan. Given Proposition 3.2, we have now dealt with all deviations that do not involve non-monitored loans.

Next, suppose that a deviant entrepreneur first chooses scale $\hat{q} = q_A^* = q_B^*$ in period 0 and then offers a non-monitored loan contract \hat{R} in period 1. We shall construct an equilibrium of the subsequent period 2 subgame, for small enough β , in which the deviant contract is rejected in period 3. In period 2, all non-deviant entrepreneurs offer the monitored loan contract $(R_G(\beta), \pi_G(\beta))$ is as defined in the proof of Proposition 3.2. In period 3, lenders accept $N_G(\beta)$ such contracts and reject the deviant's contract. In period 4, the entrepreneur announces future price $P_G(\beta)$. Given the proof of Proposition 3.2, we only have to show that it is indeed an equilibrium for the deviant contract to be rejected.

Suppose, contrahypothesis that the deviant contract is also accepted. Let \hat{P} be the equilibrium price in the subsequent period 4 subgame and let $\hat{\pi}$ be the deviant firms success probability in that same equilibrium; where $\hat{P} - \varphi'(\hat{\pi}) = \hat{R}$. Since there is only one deviant, secondary asset markets are thin. Therefore for the deviant contract to have been accepted requires that $\hat{R} \geq vk(q_A^*)/q_A^*\hat{\pi}$. Otherwise, the deviant lender would make expected losses. Combining we have $\hat{P} - \varphi'(\hat{\pi}) \geq vk(q_A^*)/q_A^*\hat{\pi}$. But, by definition, P_B^* is the smallest price for which there exists a $\hat{\pi} \in [0,1]$ such that $P - \varphi'(\hat{\pi}) \geq vk(q_A^*)/q_A^*\hat{\pi}$. And by construction, $\hat{P} < P_G(\beta)$. Condition 3.2(ii) implies that, for sufficiently small β , $P_G(\beta) < P_B^*$, hence we have a contradiction.

Similar deviations which start with $\hat{q} \neq q_B^*$, are easier to rule out, since q_B^* is the efficient scale for non-monitored loans. If a lender deviates by setting up a discount house in period 0 then it is an equilibrium for all entrepreneurs to ignore him in period 1 since any deviant would face a thin market.

We now move to the necessary condition. Consider again an entrepreneur deviating to scale q_B^* in period 0 and offering a non-monitored non-tradeable loan contract R_B^* in period 1. By Assumption A.1 and the definition of R_B^* , for small β , the deviant entrepreneur and his lender will make positive expected profit if the deviant is the only firm accepted in period 3. Indeed, if the deviant entrepreneur is accepted and the subsequent period 4 equilibrium price is strictly greater than P_B^* then both the deviant and his lender will make positive expected profit. To prevent this deviation, therefore,

we need to construct an equilibrium of the subsequent period 2 subgame in which the deviant's contract is rejected in period 3. That is, we need to find an equilibrium such that, if the deviant contract were accepted then the subsequent period 4 price would be lower than P_B^*

Consider the period 1 subgame following the deviation to q_B^* in period 0. All non-deviants have committed to a scale inefficient for non-monitored loans in period zero. That is $q_G^* > q_B^*$ and $k'' > 0$, therefore $vk(q_G^*)/q_G^* > vk(q_B^*)/q_B^*$. Therefore any equilibrium path of this subgame in which there are non-monitored firms other than the deviant active in period 4 must involve a price strictly greater than P_B^* . But then, for small enough β , the deviant's contract, R_B^* , would be accepted in period 3 since adding one additional firm would still leave the subsequent period 4 equilibrium price above P_B^* . Thus, for small β , the deviation can not be deterred using the threat of other non-monitored firms to depress the price.

Next we try to deter this deviation using monitored firms to depress the price. Consider the period 2 subgame following the deviation to q_B^* in period 0 and the offer of R_B^* in period 1. From Proposition 3.2, we know that any equilibrium path of this subgame involving active monitored firms in period 4 must involve a price weakly greater than P_G^* . Consider the price that would emerge if the deviant contract were also accepted in period 3. If $P_G^* - \beta q_B^* \pi_B^* > P_B^*$, contrary to our necessary condition, then the deviant and its lender will make positive expected profit. Hence the condition. ■

Appendix 3, 16 may 1995

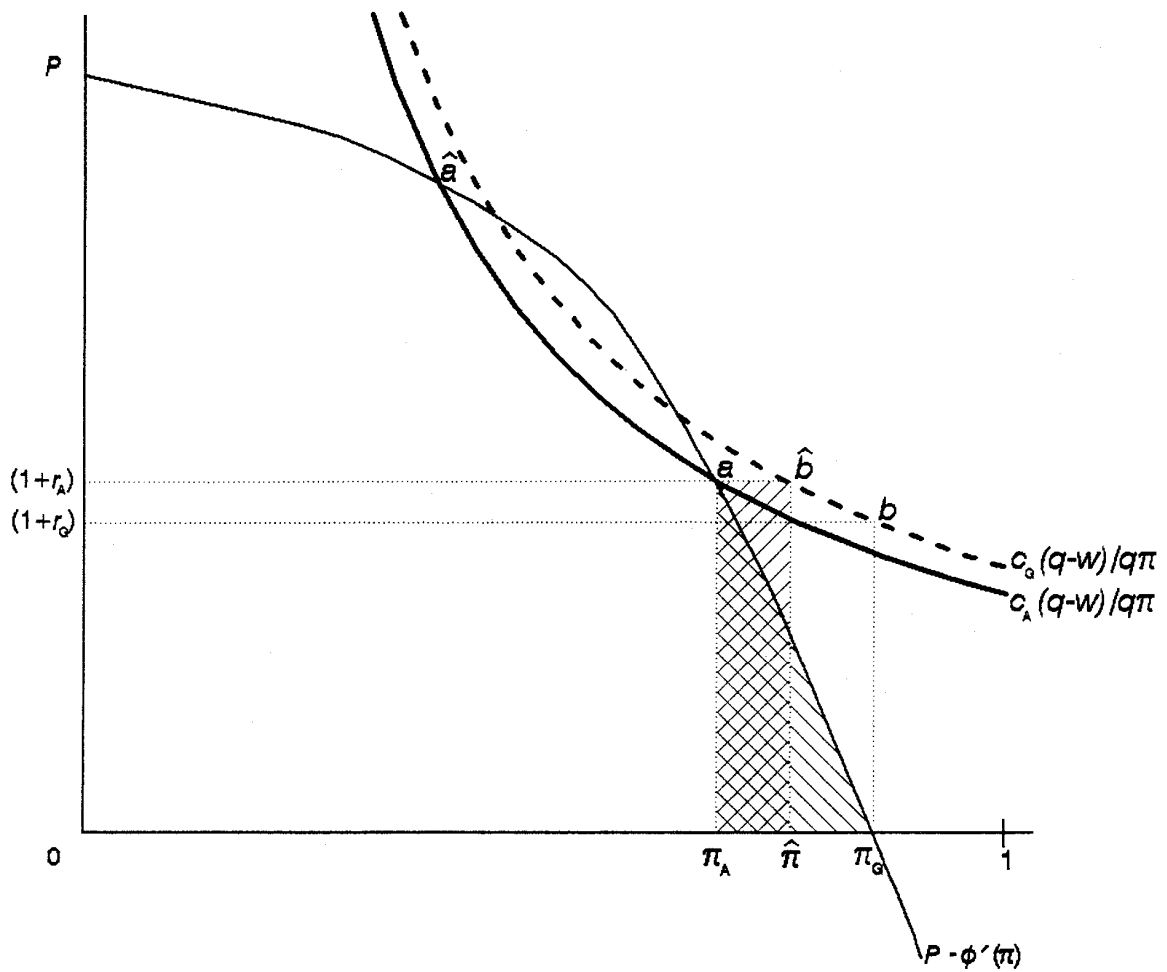


Figure 2.1 (i)

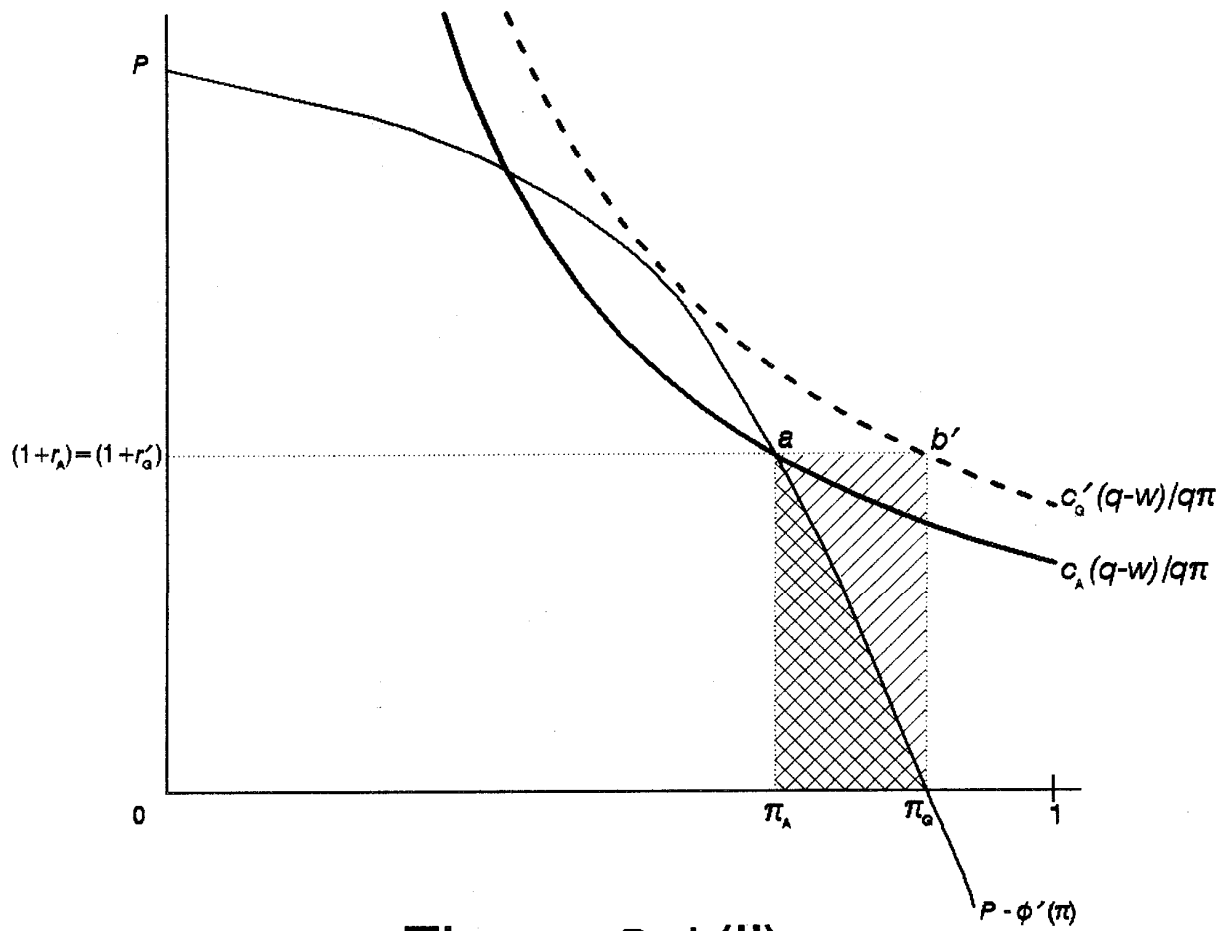


Figure 2.1 (ii)

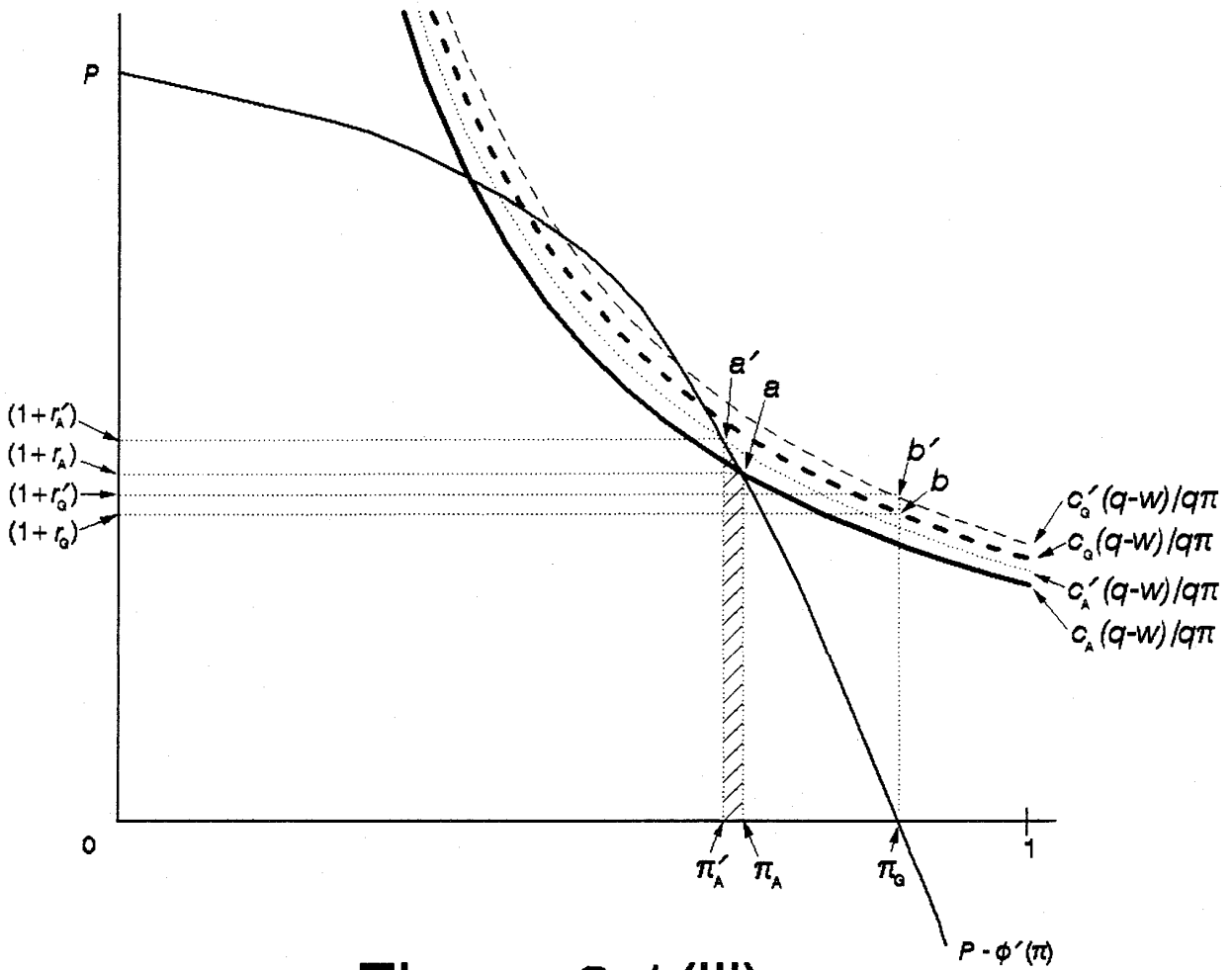


Figure 2.1 (iii)

British & German Nominal Interest Rates
1755 - 1914
Long Term Bond Yields

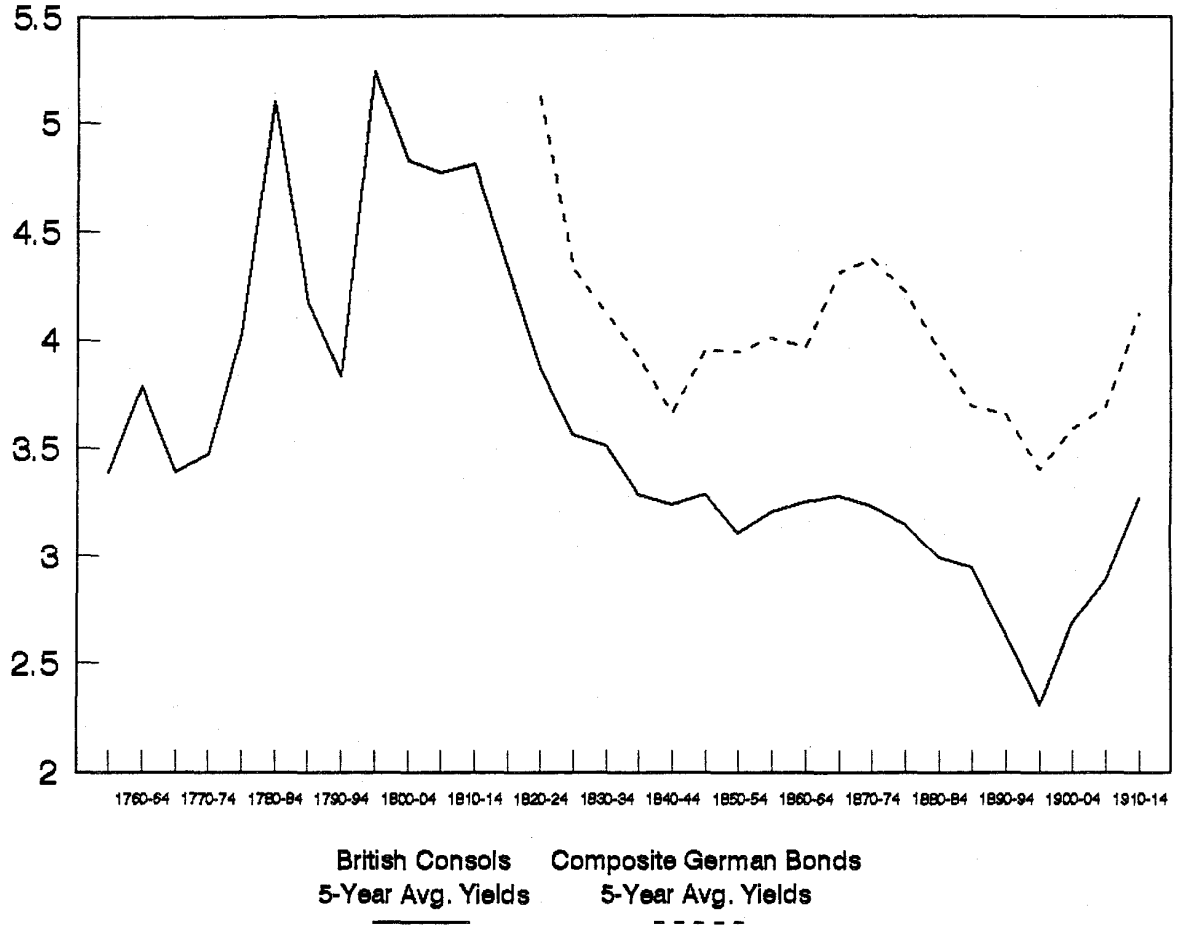


Figure 2.1 (iv)

Data Source: Homer and Sylla (1991); pp. 161-2, pp. 195-7, pp. 259-61, pp. 444-445, pp. 504-5.

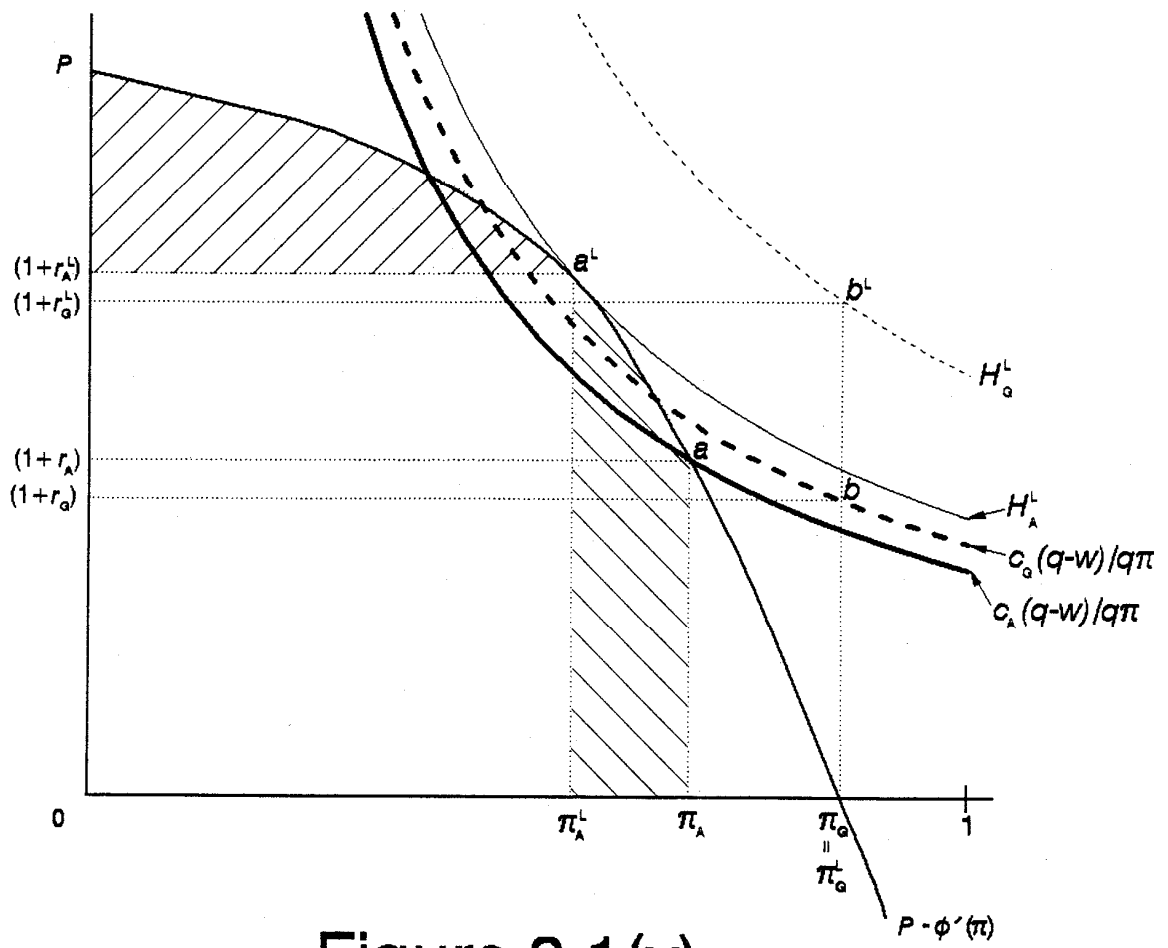


Figure 2.1(v)

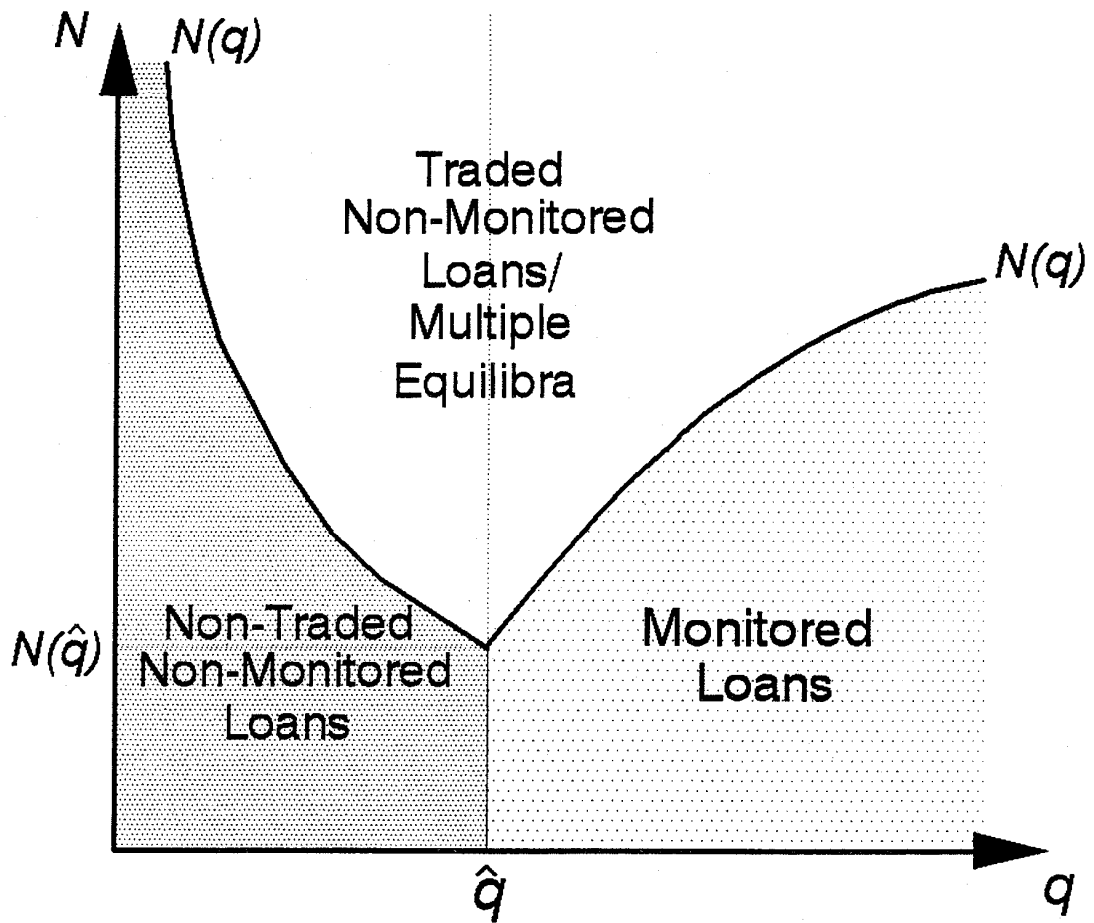


Figure 2.2(i)

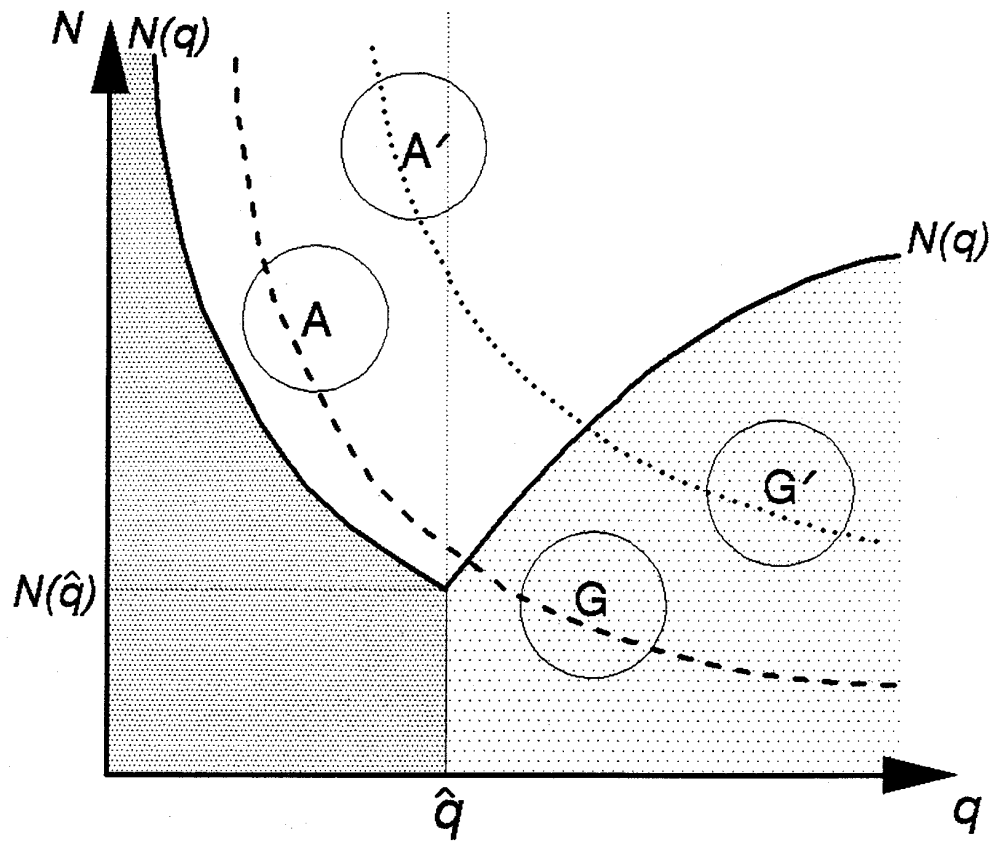
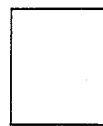
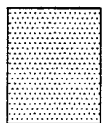


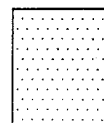
Figure 2.2(ii)



Traded
Non-Monitored Loans/
Multiple Equilibria



Non-Traded
Non-Monitored
Loans



Monitored
Loans

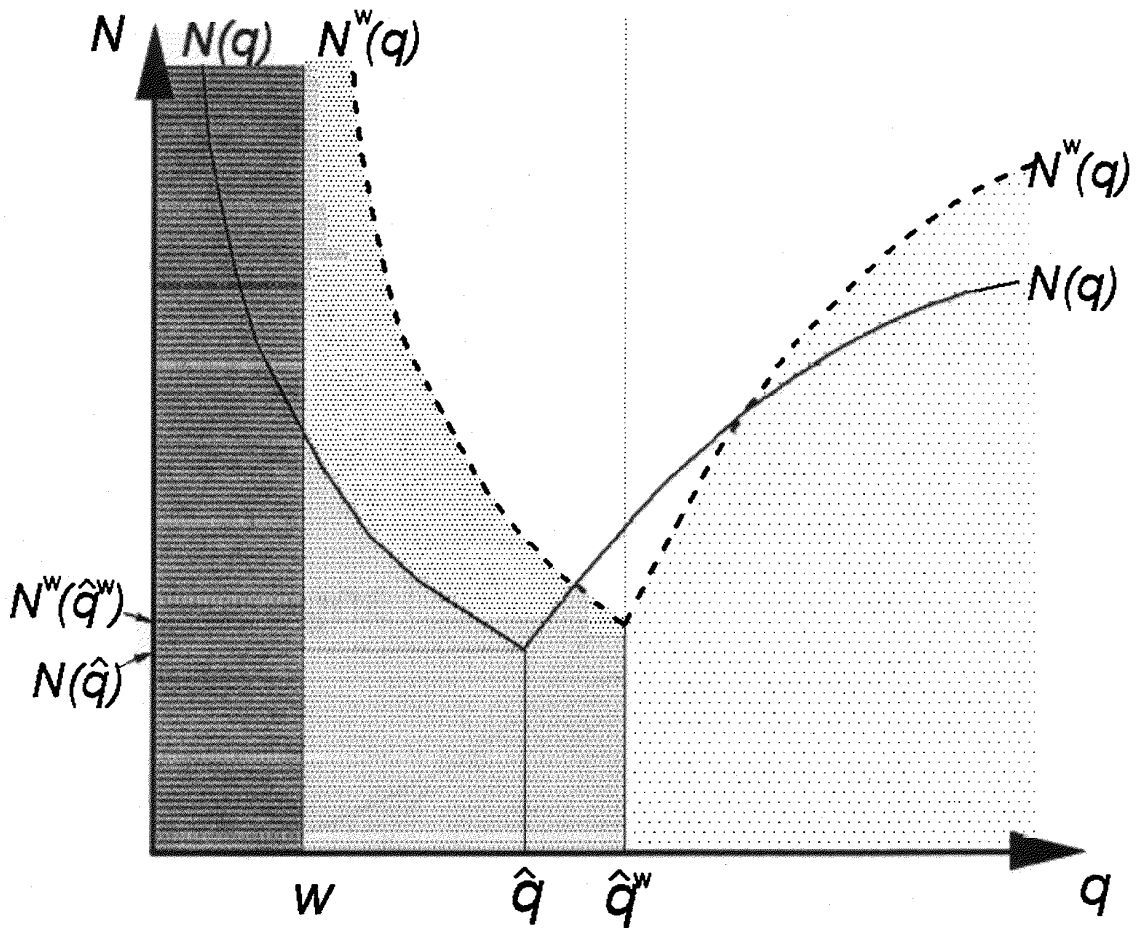
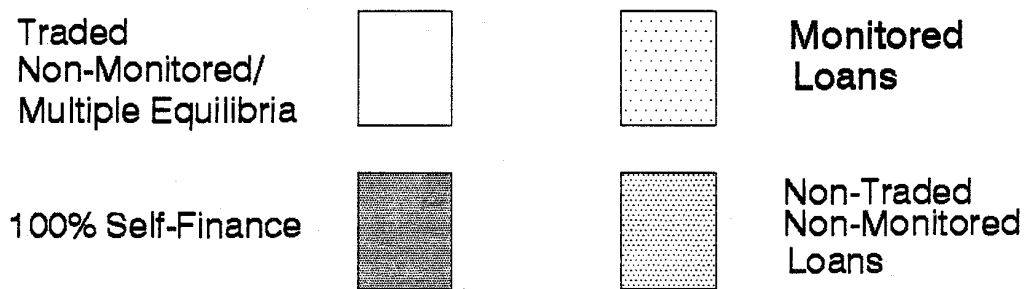


Figure 2.2(iv)



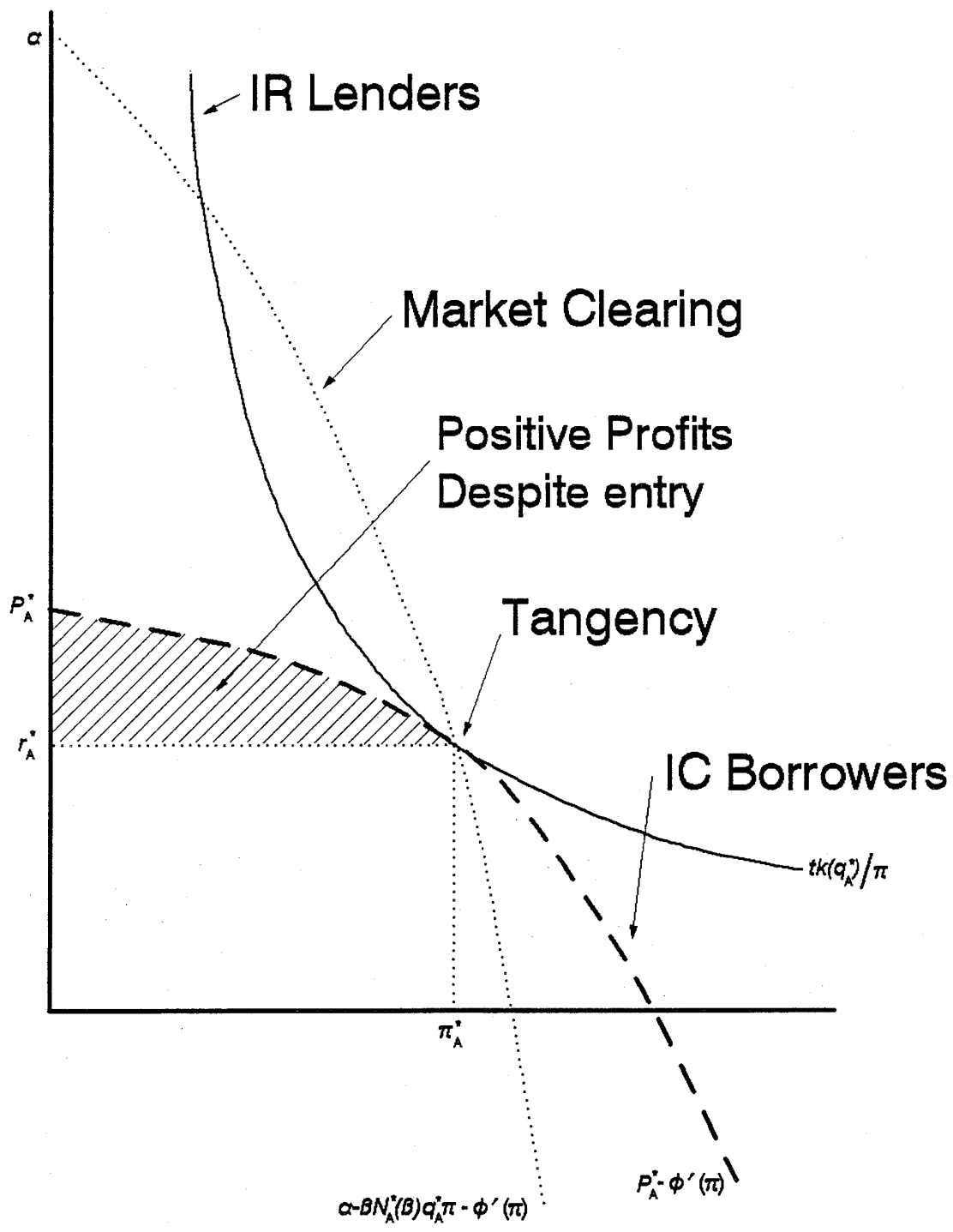


Figure 3.2(i)

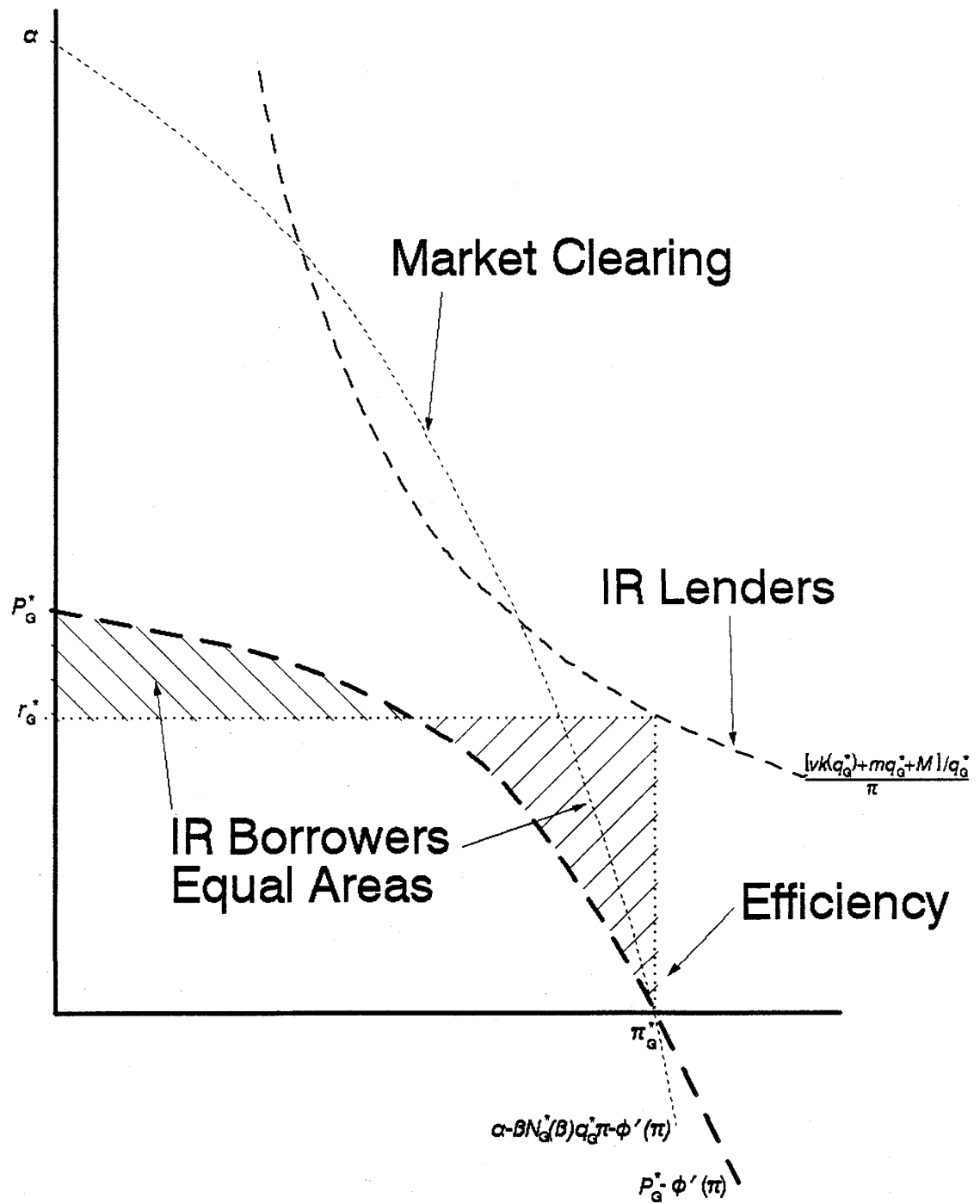


Figure 3.2(ii)

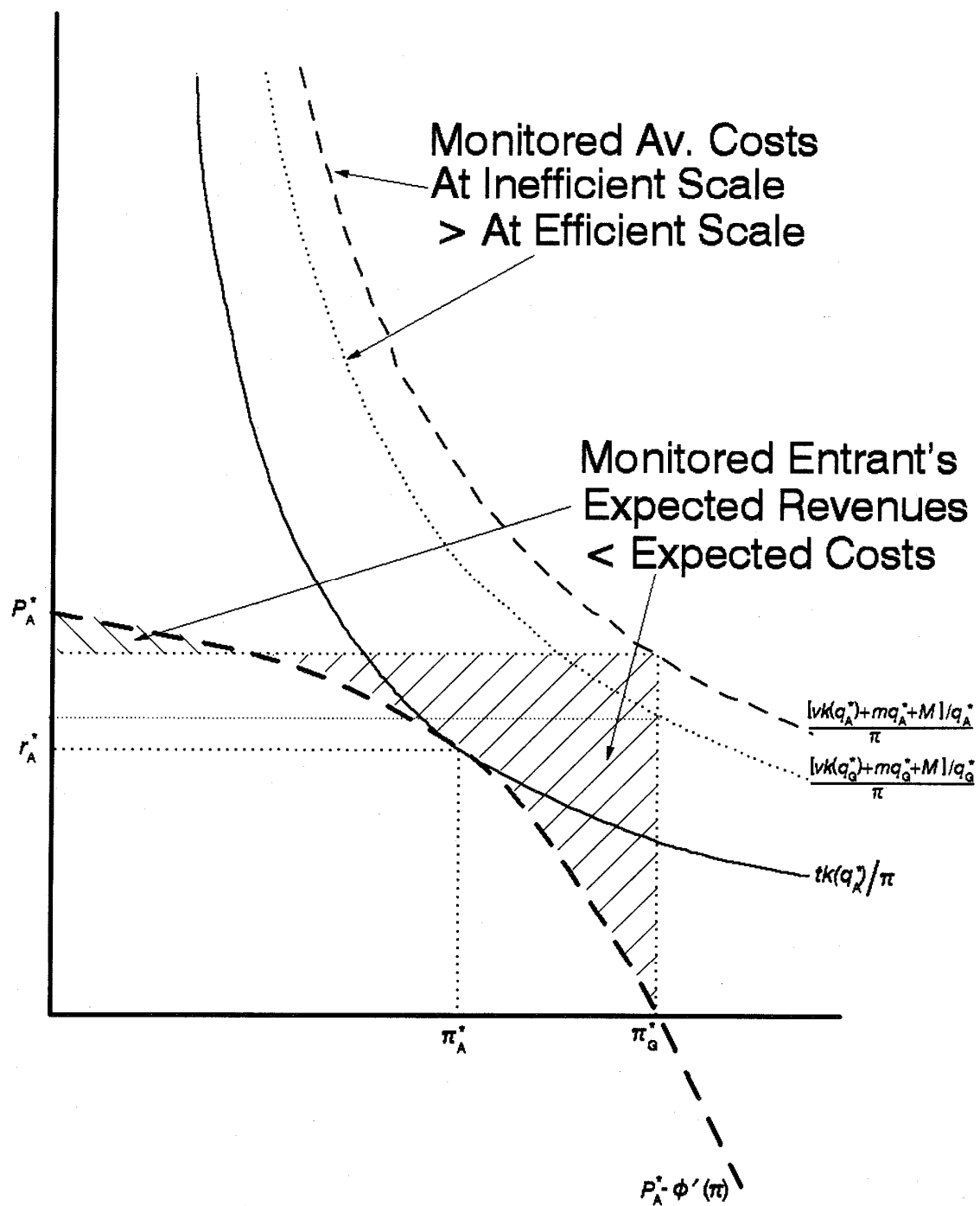


Figure 3.3(i)

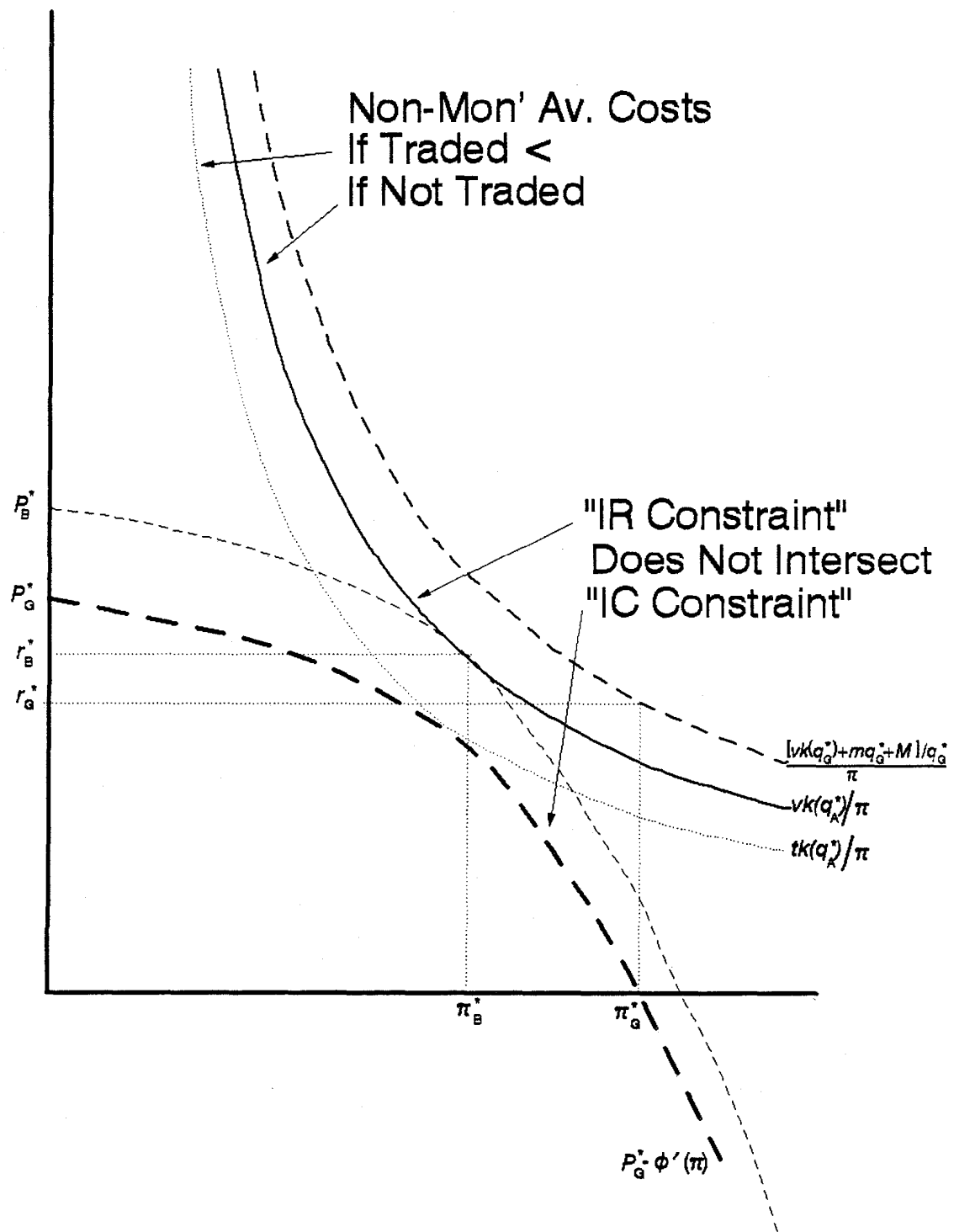


Figure 3.3(ii)

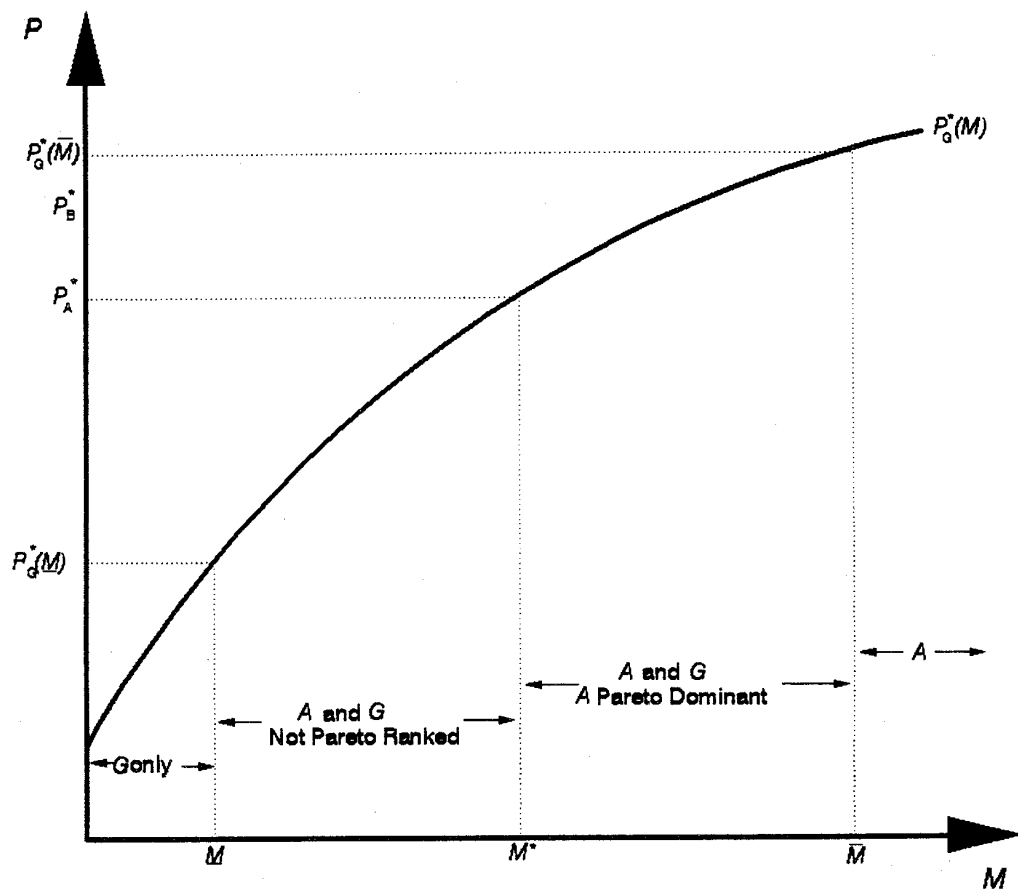


Figure 3.4