Competition and Financial Stability[1]

Franklin Allen
Department of Finance
Wharton School
University of Pennsylvania
Philadelphia, PA 19104
allenf@wharton.upenn.edu

Douglas Gale
Department of Economics
New York University
269 Mercer Street
New York, NY 10003
douglas.gale@nyu.edu

September 6, 2003

Abstract

Competition policy in the banking sector is complicated by the necessity of maintaining financial stability. Greater competition may be good for (static) efficiency, but bad for financial stability. From the point of view of welfare economics, the relevant question is: What are the efficient levels of competition and financial stability? We use a variety of models to address this question and find that different models provide different answers. The relationship between competition and stability is complex: sometimes competition increases stability. In addition, in a second-best world, concentration may be socially preferable to perfect competition and perfect stability may be socially undesirable.

1 Introduction

In the banking sector, unlike other sectors of the economy, competition policy must take account of the interaction between competition and financial stability. Greater competition may be good for (static) efficiency, but bad for financial stability.[1] In this paper, we shall argue that the relationship between competition and financial stability is considerably more complex than this simple "tradeoff" suggests, but understanding why increasing competition might reduce economic stability is a good starting point.

In an important paper, Keeley (1990) provided a theoretical framework and empirical evidence that deregulation of the banking sector in the US in the 1970's and 1980's had increased competition and led to a reduction in monopoly rents. This reduction in 'charter value' magnified the agency problem between bank owners and the government deposit insurance fund. The bank owners or managers acting on their behalf had an increased incentive to take on extra risk, given the guaranteed funds available to them because of deposit insurance. As in the agency problem identified by Jensen and Meckling (1976), if the gamble was successful the equity owners would obtain the rewards while if it was unsuccessful the cost would be born by the deposit insurance fund. The extra risk that banks took on as a result of this agency problem caused a dramatic increase in bank failures during the 1980's. The US is not the only country where there appears to be an empirical relationship between increased competition and financial instability. Beck, Demirguc-Kunt and Levine (2003) find using data from 79 countries that crises are less likely in more concentrated banking systems.

Various empirical studies have found that the cost of financial instability is high. For example, Hoggarth and Saporta (2001) find that the average fiscal costs of banking resolution across countries are 16% of GDP. For emerging countries the figure is 17.5% and for developed countries it is 12%. As Table 1 shows, the costs of banking crises alone are estimated at 4.5% of GDP. Although these costs are substantial, they are much lower than the costs, estimated at 23% of GDP, of banking and currency crises occurring together. A proportion of the fiscal costs are transfers, so these figures do not represent the deadweight economic costs. A number of studies measure the cumulative output loss resulting from a financial crisis by using the deviation from trend output. Table 2 gives estimates for these costs. The average cumulative output loss for all crises is 16.9% of GDP. Here the costs of twin crises are again higher: the loss caused by twin banking and currency crises is 29.9% of GDP versus 5.6% for banking crises alone. However, in contrast to fiscal costs, developed countries have a greater loss, 23.8% of GDP, than emerging countries, 13.9% of GDP.

The large literature on the efficiency of the banking industry (for a survey see, e.g., Berger and Humphrey (1997)) is mostly concerned with the cost- and profit-efficiency of retail banking. For example, Canoy et al. (2001) summarize the evidence as suggesting that the average bank operates at a cost level that is 10 or 20 percent above the best-practices level. This is just one (probably small) part of the total costs of deviations from perfect competition. Unfortunately, the total costs of a deviation from perfect competition have not

---

[1] See Canoy et al. (2001) and Carletti and Hartmann (2002) for excellent surveys of the literature on financial stability and competition.

1

been documented as carefully as the costs of financial instability.

Given the large and visible costs of financial instability, it is natural for policymakers to make the avoidance of financial crises a high priority. By contrast, the difficulty of measuring the efficiency costs of concentration may suggest that competition policy warrants a lower priority. In fact, the uncertainty about the costs of concentration together with the perceived (negative) tradeoff between competition and financial stability may actually encourage policymakers to favor concentration at the expense of competition policy. This subordination of competition policy to financial stability may be unwise for a number of reasons, however. In the first place, the extent to which there is a negative tradeoff between competition and financial stability may be questioned. The costs of financial crises are undoubtedly high, but it does not follow that it is necessary to reduce competition to avoid those costs. Secondly, the wide range of estimates of the efficiency costs from concentration is at least consistent with a high efficiency gain from greater competition. Thirdly, the costs of financial crises occur infrequently, perhaps every decade or few decades, whereas the inefficiecy costs concentration are borne continuously.

The proper balance between competition and financial stability presupposes a framework in which we can identify the welfare costs and benefits of different levels of competition and financial stability. Our objective in this paper is to review a number of theoretical models as a prelude to the development of a theoretical framework in which the optimal policy can be identified. From the point of view of welfare economics, the relevant question is: What are the efficient levels of competition and financial stability? We use a variety of models to address this question and find that different models provide different answers. This should not be surprising. In a second-best world, concentration may be preferred to perfect competition (Schumpeter, 1950) and perfect stability may be socially undesirable (Allen and Gale, 1998). When we consider the relationship between competition and stability, we find that the idea of a simple negative tradeoff is, again, too simple: sometimes competition decreases stability and sometimes perfect competition is compatible with the socially optimal level of stability.

We begin in Section 2 by describing the general equilibrium model of financial intermediaries and markets from Allen and Gale (2003a). Allen and Gale (2003a) provide analogues of the classical theorems of welfare economics for a model of intermediation with asymmetric information. If financial markets are complete and contracts between intermediaries and their customers are complete, the perfectly competitive equilibrium allocation is incentive-efficient. In this sense, perfect competition is socially optimal. There is no financial instability because contracts are completely contingent and hence there is no need to default. Similarly, if contracts are incomplete, the perfectly competitive equilibrium allocation is constrained-efficient, but now financial instability is necessary for efficiency. If the banks cannot meet the fixed payments that they have promised, there is a financial crisis. A deviation from competition may increase financial stability, but cannot increase and is likely to reduce welfare. This result illustrates that, in general, there should be no presumption that reducing competition in order to increase financial stability is socially desirable.

In simple partial-equilibrium models, it is possible to generate a negative trade-off between competition and financial stability. However, even in this case, the nature of the

2

trade-off between competition and stability is more complicated than was first thought. For example, Allen and Gale (2000a, Chapter 8), Boyd and De Nicolo (2002) and Perotti and Suarez (2003) have identified a number of different effects of increased competition on financial stability. In some circumstances, increased competition can actually increase financial stability. These models are discussed in Section 3.

Introducing other kinds of frictions produces further complications to our picture of competition. One the one hand, as mentioned above, Allen and Gale (2000a, Chapter 8) show that when search costs are introduced, competition among a large number of unitary banks may result in the monopoly price being charged. On the other hand, a system with two banks with branches at each location may result in the perfectly competitive price. A Hotelling-type model of spatial competition introduces a rich variety of effects concerning regional diversification and risk sharing. The profitability of banks is shown to be extremely sensitive to the precise form of local interactions.

Section 5 describes a model of Schumpeterian competition, in which firms compete by developing new products. The firm that makes the best innovation manages to capture the whole market. The equilibrium price equals the difference between the value of the successful firm's product and the value of the second-best product. So the successful firm's profit is equal to the social value of its innovation. This provides the firms with the right incentives to innovate efficiently. In this context, perfect competition is again desirable and can lead to efficiency. Clearly, this kind of innovation process is not consistent with financial stability. The successful innovator will survive while the unsuccessful will fail. Again, as in the benchmark model, efficiency requires a combination of perfect competition and financial instability. If the government is concerned with financial stability it may ensure that banks survive by regulating entry in different submarkets. We consider a setting where banks are assured of a monopoly in their region and consider the incentives to innovate. It is shown that this enforced stability leads to a welfare loss as might be expected. Less obvious is the result that there may be too little or too much investment in innovation. There can be too little because each bank only obtains profits from its own region. There can be too much because the bank is assured of a return no matter what happens.

Contagion is another important source of financial instability. It occurs when some shock, possibly small, spreads throughout the financial system and causes a systemic problem. Allen and Gale (2000b) developed a model of contagion with a perfectly competitive banking sector. It was shown that a shock that was arbitrarily small relative to the economy as a whole could cause all the banks in the financial system to go bankrupt. The contagion spreads through the interbank market. Section 6 extends the Allen and Gale (2000b) model of contagion to allow for imperfect competition in the banking sector. It is shown that in this case the economy is not as susceptible to contagion as it is with perfect competition. Each oligopolistic bank realizes that it's actions affect the price of liquidity. By providing sufficient liquidity to the market they can ensure that contagion and their own bankruptcy are avoided. In this case there is a trade-off between competition and stability.

Concluding remarks are contained in Section 7.

2 Competition and crises

In the Arrow-Debreu model of general equilibrium, the fundamental theorems of welfare economics show that perfect competition is a necessary condition for efficiency. Allen and Gale (2003a) show that analogous results hold for a model of financial crises with complete markets. In this setting, perfect competition is compatible with the efficient level of financial stability. In this sense, there is no "tradeoff" between competition and stability. We begin by describing the model of perfect competition in a financial system consisting of financial intermediaries and markets and then summarize the theoretical results in Allen and Gale (2003a).

There are three dates $t = 0, 1, 2$ and a single good at each date. The good is used for consumption and investment.

The economy is subject to two kinds of uncertainty. First, individual agents are subject to idiosyncratic preference shocks, which affect their demand for liquidity (these will be described later). Second, the entire economy is subject to aggregate shocks that affect asset returns and the cross-sectional distribution of preferences. The aggregate shocks are represented by a finite number of states of nature. All agents have a common prior probability density over the states of nature. All uncertainty is resolved at the beginning of date 1, when the aggregate state is revealed and each agent discovers her individual preference shock.

Each agent has an endowment of one unit of the good at date 0 and no endowment at dates 1 and 2. So, in order to provide consumption at dates 1 and 2, they need to invest.

There are two assets distinguished by their returns and liquidity structure. One is a short-term asset (the *short asset*), and the other is a long-term asset (the *long asset*). The short asset is represented by a storage technology: one unit invested in the short asset at date $t = 0, 1$ yields a return of one unit at date $t + 1$. The long asset yields a return after two periods. One unit of the good invested in the long asset at date 0 yields a random return of more than 1 unit of the good that depends on the aggregate state at date 2.

Investors' preferences are distinguished ex ante and ex post. At date 0 there is a finite number n of types of investors, indexed by $i = 1, ..., n$. We call i an investor's *ex ante* type. An investor's ex ante type is common knowledge and hence contractible.

While investors of a given ex ante type are identical at date 0, they receive a private, idiosyncratic, preference shock at the beginning of date 1. The date 1 preference shock is denoted by $\theta_i \in \Theta_i$, where Θ_i is a finite set. We call θ_i the investor's *ex post* type. Because θ_i is private information, contracts cannot be explicitly contingent on θ_i.

Investors only value consumption at dates 1 and 2. An investor's preferences are represented by a von Neumann-Morgenstern utility function, $u_i(c_1, c_2; \theta_i)$, where c_t denotes consumption at date $t = 1, 2$. The utility function $u_i(\cdot; \theta_i)$ is assumed to be concave, increasing, and continuous for every type θ_i. Diamond and Dybvig (1983) assumed that consumers were one of two ex post types, either early diers who valued consumption at date 1 or late diers who valued consumption at date 2. This is a special case of the preference shock θ_i. The present framework allows for much more general preference uncertainty.

Allen and Gale (2003a) consider two different versions of the model, depending on the kind

of contracts financial institutions offer to their customers. In the first version, contracts are completely contingent, subject only to incentive-compatibility constraints. More precisely, contracts are required to be incentive-compatible and are allowed to be contingent on the aggregate states η and individuals' reports of their ex post types. Each intermediary offers a single contract and each ex ante type is attracted to a different intermediary.

One can, of course, imagine a world in which a single "universal" intermediary offers contracts to all ex ante types of investors. A universal intermediary could act as a central planner and implement the incentive-efficient allocation of risk. There would be no reason to resort to markets at all. Our world view is based on the assumption that transaction costs preclude this kind of centralized solution and that decentralized intermediaries are restricted in the number of different contracts they can offer. This assumption provides a role for financial markets in which financial intermediaries can share risk and obtain liquidity.

At the same time, financial markets alone will not suffice to achieve optimal risk sharing. Because individual economic agents have private information, markets for individual risks are incomplete. The markets that are available will not achieve an incentive-efficient allocation of risk. Intermediaries, by contrast, can offer individuals incentive-compatible contracts and improve on the risk sharing provided by the market.

In the Diamond and Dybvig (1983) model, all investors are ex ante identical. Consequently, a single representative bank can provide complete risk sharing and there is no need for markets to provide cross-sectional risk sharing across banks. Allen and Gale (1994) showed that differences in risk and liquidity preferences can be crucial in explaining asset prices. This is another reason for allowing for ex ante heterogeneity.

In the context of intermediaries with complete markets and complete contingent incentive-compatible contracts, Allen and Gale (2003a) prove the following result.

Proposition 1 *Under the maintained assumptions, the equilibrium allocation of the model with complete markets and incomplete contracts is constrained-efficient.*

Proposition 1 assumes that intermediaries use complete, incentive-compatible contracts. In reality, we do not observe such complex contracts, for reasons that are well documented in the literature, including transaction costs, asymmetric information and the nature of the legal system. These frictions can justify the use of debt and many other kinds of incomplete contracts that intermediaries use in practice. The second version of the model presented by Allen and Gale (2003a) assumes that intermediaries are restricted to using a set of incompletely contingent contracts. This framework allows for many special cases, including at one extreme the earlier model with completely contingent contracts and completely noncontingent debt contracts. Note that this framework allows for a wide variety of assumptions about what is feasible, but takes the set of feasible contracts as given. To endogenize the set of feasible contracts one would have to appeal to factors such as transaction costs, nonverifiable information, and so on.

When contracts are complete, there is no incentive for intermediaries to enter into commitments that they cannot carry out. When contracts are constrained to be incomplete, it may be (ex ante) optimal for the intermediary to plan to default in some states. In the

5

event of default, it is assumed that the intermediary's assets, including the Arrow securities it holds, are liquidated and the proceeds distributed among the intermediary's investors. For markets to be complete, which is an assumption we maintain here, the Arrow securities that the bank issues must be default free. Hence, we assume that these securities are collateralized and their holders have priority. Anything that is left after the Arrow security holders have been paid off is paid out pro rata to the depositors. Allen and Gale (2003a) demonstrate the following result for the case when contracts are incomplete and take the form of deposit contracts.

Proposition 2 *Under maintained assumptions, the equilibrium allocation of the model with complete markets and incomplete contracts is constrained-efficient.*

This is an important result. It shows that, in the presence of complete markets and perfect competition, the incidence of default is optimal in a laisser-faire equilibrium. There is no scope for *welfare-improving* government intervention to prevent financial crises. In fact competition and financial instability are both necessary for constrained efficiency.

This result demonstrates that in a standard framework achieving optimality does not require trading off competition and financial stability. As we will see below this result extends to a number of other circumstances.

It is important to stress that the results in this section are simply benchmarks to illustrate what may happen. The only costs modelled are the losses to consumers from inefficient risk sharing. Many features that may be important in practice, such as unemployment and bankruptcy costs to firms are excluded. When these are taken into account, there may be a role for government intervention to reduce the incidence of financial crises. What the results do show is that the operation of the financial system and the occurrence of crises when there are complete markets are not the problem. There must be some form of market failure for financial crises to be undesirable.

3   Agency costs

Keeley (1990) developed a simple model of risk taking by banks with two dates and two states when there is deposit insurance. He showed that as competition increased risk taking by banks also increased. In fact deposit insurance is not necessary for this effect to be present although it does exacerbate it. Allen and Gale (2000a, Chapter 8) develop a simple model of competition and risk taking to illustrate the agency problem.

When firms are debt-financed, managers acting in the shareholders' interests have an incentive to take excessive risks, because the debtholders bear the downside risk while the shareholders benefit from the upside potential. This well known problem of risk shifting is particularly acute in the banking sector where a large proportion of the liabilities are in the form of debt (deposits). The risk-shifting problem is exacerbated by competition. Other things being equal, greater competition reduces the profits or quasi-rents available to managers and/or shareholders. As a result, the gains from taking excessive risks become

relatively more attractive and this increases the incentive to exploit the non-convexity in the payoff function. Any analysis of the costs and benefits of competition has to weigh this effect against the supposed efficiency gains of greater competition.

To illustrate these ideas, consider the problem faced by a banking regulator who controls entry into the banking industry by granting charters to a limited number of banks. We use a model of Cournot competition, in which banks choose the volume of deposits they want, subject to an upward sloping supply of funds schedule. Having more banks will tend to raise the equilibrium deposit rate and increase the tendency to shift risks. What is the optimal number of charters? Should the regulator restrict competition by granting only a few charters or encourage competition by granting many?

3.1   A Static Model

Suppose that the regulator has chartered n banks, indexed $i = 1, ..., n$.

Each bank chooses a portfolio consisting of perfectly correlated risks. This assumption is equivalent to assuming that the risk of each investment can be decomposed into a common component and a purely idiosyncratic component. If there is a very large number of investments, the purely idiosyncratic components can be pooled perfectly. Then the idiosyncratic risks disappear from the analysis and we are left with a common component representing the systematic risks.

A portfolio is characterized by its size and rate of return. The bank's investments have a two-point return structure: for each dollar invested, bank i will receive a return y_i with probability $p(y_i)$; with probability $(1 - p(y_i))$ they pay a return 0. The bank chooses the riskiness of its portfolio by choosing the target return y_i on its investments. The function $p(y_i)$ is assumed to be twice continuously differentiable and satisfies

$$p(0) = 1, p(\bar{y}) = 0, \text{ and } p'(y_i) < 0, p''(y_i) \leq 0, \forall 0 < y_i < \bar{y}.$$

The higher the target return, the lower the probability of success and the more rapidly the probability of success falls. Because the investments have perfectly correlated returns, the portfolio return has the same distribution as the returns to the individual investments.

Let $d_i \geq 0$ denote the total deposits of bank i, which is by definition the total number of dollars the bank has to invest. (For the moment, we ignore bank capital). There is an upward sloping supply-of-funds curve. If the total demand for deposits is $D = \sum_i d_i$, then the opportunity cost of funds is $R(D)$, where $R(D)$ is assumed to be a differentiable function satisfying

$$R'(D) > 0, R''(D) > 0, R(0) = 0 \text{ and } R(\infty) = \infty.$$

We assume that all deposits are insured, so the supply of funds is independent of the riskiness of the banks' portfolios. In the sequel, we consider the case where banks bear the cost of deposit insurance.

The payoff to bank i is a function of the riskiness of its own portfolio and the demand for deposits of all the banks:

$$\pi_i(\mathbf{y}, \mathbf{d}) = p(y_i)[y_i d_i - R(D)d_i],$$

where $\mathbf{d} = (d_1, ..., d_n)$ and $\mathbf{y} = (y_1, ..., y_n)$. Note that we have ignored the cost of deposit insurance to the bank in calculating its net return.

Since a bank can always ensure non-negative profits by choosing $d_i = 0$, it will always earn a non-negative expected return in equilibrium, that is, $y_i d_i - R(D)d_i \geq 0$. There is no need to introduce a separate limited-liability constraint.

In a Nash-Cournot equilibrium, each bank i chooses an ordered pair (y_i, d_i) that is a best response to the strategies of all the other banks. Consider an equilibrium $(\mathbf{y}, \mathbf{d})$ in which each bank i chooses a strictly positive pair $(y_i, d_i) \gg 0$. As a necessary condition for a best response, this pair must satisfy the following first-order conditions:

$$p(y_i)[y_i - R(D) - R'(D)d_i] = 0$$

$$p'(y_i)[y_i - R(D)]d_i + p(y_i)d_i = 0.$$

Assuming the equilibrium is symmetric, that is, $(y_i, d_i) = (y, d)$ for every i, the first-order conditions reduce to

$$y - R(nd) - R'(nd)d = 0$$

$$p'(y)[y - R(nd)] + p(y) = 0$$

and this implies that

$$-\frac{p(y)}{p'(y)} = y - R(nd) = R'(nd)d.$$

Given our assumptions on $p(y)$, an increase in y reduces $-p(y)/p'(y)$. Suppose that there are two symmetric equilibria, (y, d) and (y', d'). Then $y > y'$ implies that $R'(nd)d < R'(nd')d'$, which, given our assumptions on $R(D)$, implies that $d < d'$ and $R(nd) < R(nd')$. Then clearly $y - R(nd) > y' - R(nd')$, contradicting the first equation. So there is at most one solution to this set of equations, which determines both the size and the riskiness of the banks' portfolio in a symmetric equilibrium.

Proposition 3 *Under the maintained assumptions, there is at most one symmetric equilibrium $(y^*, d^*) \gg 0$ which is completely characterized by the conditions*

$$-\frac{p(y)}{p'(y)} = y - R(nd) = R'(nd)d.$$

What can we now say about the effect of competition on risk taking?

Suppose that we identify the degree of competitiveness of the banking sector with the degree of concentration. In other words, the larger the number of banks, the more competitive the banking sector is. So a first attempt at answering the question would involve increasing n ceteris paribus and observing how the riskiness of the banks' behavior changes.

With a fixed supply-of-funds schedule $R(\cdot)$ it is most likely that the volume of deposits will remain bounded as n increases. More precisely, if we assume that $R(D) \to \infty$ as $D \to \infty$ then it is clear that $D \to \infty$ is inconsistent with equilibrium. Then the equilibrium value of D is bounded above (uniformly in n) and this implies that $d \equiv D/n \to 0$ as $n \to \infty$. This in turn implies that $R'(nd)d \to 0$ from which it immediately follows that $y - R(nd) \to 0$ and $p(y) \to 0$, or in other words, that y and $R(nd)$ both converge to $\bar{y}$.

Proposition 4 *If $R(D) \to \infty$ as $D \to \infty$ then in any symmetric equilibrium, $y - R(nd) \to 0$ and $y \to \bar{y}$ as $n \to \infty$.*

The effect of increasing competition is to make each bank much smaller relative to the market for funds and this in turn reduces the importance of the price effect (the $R'(nd)d$ term) in the bank's decision. As a result, banks behave more like perfect competitors and will increase their business as long as profits are positive. Equilibrium then requires that profits converge to zero, and this in turn implies that banks have extreme incentives for risk taking. In the limit as $n \to \infty$, they will choose the riskiest investments possible in an attempt to earn a positive profit.

3.1.1 The Effect of Replicating the Market

This exercise is enlightening but somewhat artificial since it assumes that we are dealing with a market of fixed size and increasing the number of banks without bound in order to achieve competition. Normally, one thinks of perfect competition as arising in the limit as the number of banks and consumers grows without bound. One way to do this is to *replicate* the market by shifting the supply-of-funds function as we increase the number of banks. Precisely, suppose that the rate of return on deposits is a function of the deposits *per bank*:

$$R = R(D/n).$$

In effect, we are assuming that, as the number of banks is increased, the number of depositors is increased proportionately, so that the supply of funds in relation to a particular bank is unchanged.

The effect of this change in the model is to make it more like the traditional model of a market in which, as the number of firms increases, the effect of any firm supply on the price of the product becomes vanishingly small. Here the effect of any bank's demand for deposits on the equilibrium deposit rate becomes vanishingly small in the limit as the number of banks becomes unboundedly large. To see this, note that the first-order conditions become:

$$y - R(d) - R'(d)dn^{-1} = 0$$

and

$$p'(y)[y - R(d)] + p(y) = 0.$$

As before, we can ensure that d remains bounded as $n \to \infty$ by assuming that $R(d) \to \infty$ as $n \to \infty$. Then the last term on the left hand side of the first equation will vanish as $n \to \infty$, leaving a limiting value of (y, d) that satisfies $y = R(d)$. Substituting this in the second equation tells us that $p(y) = 0$. In other words, as the number of banks increases, the profit margins fall to zero, with the result that banks choose riskier and riskier investments.

Proposition 5 *If $R(D) \to \infty$ as $D \to \infty$ then in any symmetric equilibrium, $y - R(d) \to 0$ and $y \to \bar{y}$ as $n \to \infty$.*

9

This is a highly stylized model, so the results have to be taken with a grain of salt; nonetheless, they illustrate clearly the operative principle, which is that competition, by reducing profits, encourages risk taking.

In this particular case, we have constant returns to scale in banking, so that in the limit, when there is a large number of individually insignificant banks, profits must converge to zero. In other words, banks will expand the volume of their deposits and loans until the deposit rate approaches the expected return on investments. But this gives them an extreme incentive to shift risks to the depositors or the deposit insurance agency, since it is only by doing so that they can get positive profits at all.

With constant returns to scale, zero profits is always a necessary condition of equilibrium in a competitive industry. However, there are other ways of ensuring the same outcome even if constant returns to scale is not assumed. We replicated the market by increasing the number of banks and potential depositors in the same proportion. This is an interesting thought experiment, but it is not the same as the comparative static exercise the regulator is undertaking. Presumably, the regulator has to choose n optimally, taking as given the supply-of-funds schedule. Suppose that m is the number of depositors and n the number of banks. Then the market supply-of-funds schedule can be written $R(D/m)$ if $R(\cdot)$ is the individual supply-of-funds schedule. When m is very large, the supply of funds is elastic, other things being equal, so the banks will take the marginal cost of funds as being equal to the average cost $R(D/m)$. However, increasing the number of banks n in relation to m will force profits down. If d remains bounded away from zero, the average cost of funds must increase to ∞ and if d goes to zero, profits will also go to zero. In this way, the regulator can achieve the effects of free entry, but there is no need to do this in order to ensure competition. Competition, in the sense of price-taking behavior, follows from having a large market, that is, a large value of m, independently of whether n is large or not. Clearly, the regulator does not want to drive profits to zero if it can be helped, because of the incentives for risk taking that that creates.

3.1.2   Cost of Deposit Insurance

The preceding analysis assumes that all deposits are insured and that the costs are not born by the banks. This is clearly unrealistic, so it makes sense to consider explicitly the cost of deposit insurance. We assume that the premium for deposit insurance is set before the banks choose their strategies and that it is the same for each bank, independently of the strategy chosen. In equilibrium, the premium accurately reflects the cost of deposit insurance provided by a risk neutral insurer.

Let π denote the premium per dollar of deposits. Then the objective function of bank i is $p(y_i)(y_i - R(D) - \pi)d_i$ and the first-order conditions in a symmetric equilibrium in which banks choose the strategy (y, d) will be:

$$y - R(d) - \pi - R'(d)dn^{-1} = 0$$

$$p'(y)[y - R(d) - \pi] + p(y) = 0.$$

In equilibrium, the premium must be set so that the expected return on deposits is equal to the return demanded by depositors:

$$R(d) = p(y)(R(d) + \pi).$$

Substituting $\pi = [(1 - p(y))/p(y)]R(d)$ into the first-order conditions yields

$$y - R(d)/p(y) - R'(d)dn^{-1} = 0$$

$$p'(y)[y - R(d)/p(y)] + p(y) = 0.$$

Let (y^n, d^n) be a symmetric equilibrium when there are n banks and suppose that $(y^n, d^n) \to (y^0, d^0)$ as $n \to \infty$. Then the first-order conditions imply that

$$\lim_{n \to \infty} y^n - R(d^n)/p(y^n) = 0$$

which is only possible if $y^n \to \bar{y}$ and $p(y^n) \to 0$, as before.

3.1.3   Efficiency

Let us leave distributional questions on one side for the moment, although historically they have been at the center of the arguments for competition in banking, and suppose that the regulator is only interested in maximizing surplus. Reverting to the constant-returns-to-scale case, two necessary conditions for Pareto optimality are that the average cost of funds be equal to the expected return on investments, and that the expected return on investments should be a maximum:

$$
\begin{aligned}
R(D/m) &= p(y)y \\
p(y)y &\geq p(y')y', \forall y' \in [0, \bar{y}].
\end{aligned}
$$

Neither of these conditions will hold in equilibrium when m and n are very large. The first condition requires that the volume of deposits expand until the cost of funds equals the expected value of investments. However, when the market is highly competitive, we have $y - R(D/m) \cong 0$, so that $p(y)y - R(D/m) < 0$. As we also saw, the second condition cannot be satisfied in equilibrium, since as the market grows large $(m, n \to \infty)$, we have $y \to \bar{y} < \infty$ and $p(y) \to 0$, so $p(y)y \to 0$. This is not only sub-optimal but the worst possible outcome because it minimizes the total surplus.

Suppose instead that we hold the value of m fixed and adjust n to maximize total surplus, taking the equilibrium values $(y(n), d(n))$ as given functions of n determined by the equilibrium conditions

$$y - R(nd/m) - R'(nd/m)dm^{-1} = 0$$

and

$$p'(y)[y - R(nd/m)] + p(y) = 0.$$

11

A "small change" in n will increase the expected revenue by

$$\{p'(y(n))y(n) + p(y(n))\}ny'(n) + p(y(n))y(n)$$

and the cost by

$$R(nd(n)/m)\{nd'(n) + d(n)\},$$

so a necessary condition for an (interior) optimum is

$$\{p'(y(n))y(n) + p(y(n))\}ny'(n) + p(y(n))y(n) = R(nd(n)/m)\{nd'(n) + d(n)\}.$$

This can be rewritten as

$$p(y(n))y(n) - R(nd(n)/m)d(n) = -\{p'(y(n))y(n) + p(y(n))\}ny'(n) + R(nd(n)/m)nd'(n)$$

where the left hand side is the expected surplus generated by a single bank and the right hand side is the change in expected revenue per bank as n increases plus the change in the cost per bank of the funds borrowed. Since the left hand side is positive, the right hand side must be positive, too. But we know that the second term on the right must be negative since adding more banks reduces the volume of business each bank does, so the first term on the right is positive. We know that $y'(n)$ is negative—increased competition leads to increased risk taking—so the term in braces must be positive. Assuming that $p(y)y$ is concave in y, this tells us that n will be chosen so that $y(n)$ is less than the value that maximizes expected revenue.

A fortiori, it will not, as we have seen, be as great as the value under free entry, since it is never optimal to let $n \to \infty$. It may in fact, be optimal to let the number of banks remain quite small.

3.2   Loan Market Competition

The model in Allen and Gale (2000a) analyzes competition in the deposit market. The bank is assumed to invest deposits directly in a portfolio of assets with given risk characteristics. The bank directly determines the riskiness of its portfolio. As profits decline the bank's preference for risk increases, so increasing competition leads to increasing risk and a decrease in stability. Boyd and De Nicolo (2002) point out that the assumption that banks invest directly in assets is crucial for the result. To show this, they extend the Allen and Gale model to include entrepreneurs. The entrepreneurs obtain loans from the banks and invest the money in risky ventures. Each entrepreneur chooses the riskiness of the venture he invests in. The entrepreneur's, like the banks in the Allen and Gale model, have a greater incentive to take risk when profits are lower. However, the effect of competition among banks here is the opposite of what we observed in the Allen and Gale (2000a) model. Greater competition among banks reduces the interest rates that borrowers pay, increases the profitability of their ventures and hence reduces the incentive to take risk. Thus, increased competition among banks leads to increased financial stability. The effect of competition in the deposit market

is the same as before but Boyd and De Nicolo are able to show that the loan market effect dominates. The tradeoff between competition and stability presented in the Allen and Gale model is reversed in the Boyd and De Nicolo model. As competition between banks increases the risks taken by borrowers is unambiguously reduced and financial stability is improved.

3.3   Dynamic Competition

The results in Subsection 3.1 demonstrate how the limited liability of managers and shareholders in a modern banking corporation can produce a convex objective function which in turn leads to risk-shifting behavior. This kind of behavior is most likely to occur when the bank is "close to the water line", that is, when the risk of bankruptcy is imminent. For banks which are not in immediate danger of bankruptcy, the risk-shifting argument may be less relevant. However, even if a bank is not close to the water line, there may be other reasons for thinking that its objective function is convex. Consider, for example, the winner-takes-all nature of competition. When banks compete for market share, the bank that ends up with the largest share may be able to exploit its market power to increase profitability. In this case, the profit function may be convex in market share, that is, doubling market share may more than double profits. Another reason is the presence of increasing returns to scale. If larger banks have lower average costs, then profits will be a convex function of the size of the bank. Either of these possibilities will give the bank an incentive to take riskier actions, even when the bank is not in immediate danger of bankruptcy.

These incentives for risk-taking behavior are naturally studied in a dynamic context. Suppose that a group of banks are competing over time. Their activities are constrained by the minimum capital ratio, so the only way to expand is to acquire more capital. Because of agency costs or the adverse signaling effects, it may be expensive for banks to raise capital from external sources, so they try to accumulate capital by retaining earnings. The relative size of the bank matters, because it gives the bank a competitive edge over other banks. Because their reduced-form profit functions are convex and they are constrained by their capital, the game is a race to see who can accumulate capital or market share fastest.

When banks compete to capture greater market share or to reap economies of scale, they consider not only the effect of their actions on immediate profits but also on their future position in the market. How the bank's current actions will affect its future position in the market depends on the nature of the risks involved and on the behavior of the other banks. Even if the profit function is only convex when the bank is close to the bankruptcy point, the bank's objective function may be convex over a much wider region because the bank's objective function incorporates or discounts future possibilities which are still far away. This may influence the shape of the bank's objective function globally, through backward induction.

Allen and Gale (2002a, Chapter 8) consider a variety of different models and show that risk taking can either be increased or decreased by competition in a dynamic setting. The first case analyzed involves a pair of duopolists who compete for market share. They play repeatedly for many periods. In each round market shares can go up or down a small amount.

By taking a risky action instead of a safe action in any period they increase the variance of the change in market share. A crucial assumption is that there are "reflecting barriers" at the extreme values of market share. When a bank's market share hits zero, it does not go out of business; at worst it will remain at zero for some periods before bouncing back. This non-convexity is like having increasing returns in the neighborhood of zero. Similarly, when a bank's market share hits 100% it must eventually bounce back, and this is like having locally decreasing returns. A simple numerical example is used to illustrate the effect of non-convexity on the bank's behavior. When a bank's market share is low, its objective function is convex and it has an incentive to take risk; when its market share is high, its objective function is concave and it has an incentive to avoid risk.

The second case analyzed assumes there are "absorbing barriers". Now when market share hits 0 or 1 it stays the same with probability one. In this case it is shown that the incentive to take risks is eliminated. The reason is the assumption that the bank's position can only change a little bit at a time. If the period length and the step size are made very small, the binomial process considered would approximate Brownian motion, which has continuous sample paths with probability one. It is the continuity of the movement of the bank's market share over time which eliminates the usual incentive for risk taking. The bank becomes "bankrupt" as soon as its market share hits zero. It cannot go below the line and so it cannot shift risk to depositors or other creditors. Whether incentives to take risks are greater or less than with reflecting barriers is ambiguous. While absorbing barriers eliminate the positive incentive for risk taking with reflecting barriers, they also eliminate the incentive to avoid risk when market share is high that was found in the same model.

Perotti and Suarez (2003) consider the effect of dynamic competition in a model where there can be a banking duopoly or monopoly. If a bank fails when there is a duopoly the market structure switches temporarily to a monopoly. Banks can lend prudently in which case their portfolio of loans has a safe payoff. The alternative is to lend speculatively in which case there is some probability of a high payoff but the average payoff is low. Limited liability and deposit insurance mean lending speculatively can shift risks and be advantageous in the short run. However, a bank can be hit by a random *solvency shock.* If it lent prudently it will always survive this shock but if it lent speculatively it will fail unless loan returns are high. It is shown that this effect introduces an incentive for banks to lend prudently. A prudent bank will be less exposed to the risk of being driven out of business and will emerge as a monopolist if the other duopolist lent speculatively and is hit by a solvency shock. In their model this "last one standing effect" makes duopolistic banks unambiguously more prudent and encourages stability.

The range of results obtained with these dynamic models illustrate how crucial the particular details of the model are in determining whether or not competition leads to more or less financial stability.

4 Spatial competition

So far, we have only considered competition in markets for homogeneous commodities or services. Product differentiation occurs in the financial sector, just as it does in non-financial sectors, and it is important to consider the effect of product differentiation on the competitive process. Models of spatial competition are used to represent competition among firms with differentiated products and the same can be done for the banking sector. We can interpret the spatial dimension literally as representing banks with different locations or we can interpret it metaphorically as representing some other qualitative difference in the services provided. In either case, we find that the effects of concentration on competition in spatial models can be quite different from the results obtained for markets with undifferentiated products. In this section we consider two models of bank competition. The first model, from Allen and Gale (2000a), compares branch banking with unitary banking and shows that competition among a small number of banks (with many branches) may be more aggressive than competition among a large number of unitary banks. We then consider a Hotelling-type model of spatial competition and again compare competition among a small number of banks (with many branches) with competition among a large number of unitary banks. The results concerning the trade-off between competition and diversification are very sensitive to the spatial arrangement of the branches.[2]

4.1 Unitary Banking versus Branch Banking

The model presented below exploits two types of imperfections arising from asymmetric information. The first is the presence of "lock-in" effects. Information is costly for both banks and their customers, whether borrowers or depositors, and once relationship-specific investments in information have been made, the parties may find themselves "locked-in" to the relationship. For example, a borrower having incurred a fixed cost of revealing its type to a bank, will suffer a loss if it switches to another bank. In addition there is the well known "lemons effect", that arises if the borrower leaves the bank with which it has been doing business for many years. These lock-in effects will be modeled by simply assuming that there is a fixed cost of switching banks. As is well known (see, e.g., Diamond (1971)), switching costs give the bank a degree of monopoly power, even if the bank is not "large".

There are other reasons why banks are monopolistic competitors. For locational reasons their services are not perfect substitutes. Differences in size and products and specialized knowledge also make them imperfect substitutes. Here, we focus on the lock-in effect.

The second essential imperfection arises from the fact that a bank's customers have incomplete information about the services offered by a bank and the prices at which these services are offered, at the time when the relationship is begun. In fact, the smaller the bank is, the less likely it is that the bank's reputation will be an adequate source of information about the quality and prices of the bank's products.

---

[2]For other welfare analyses involving spatial competition and risk see Besanko and Thakor (1992) and Matutes and Vives (1996, 2000).

A third important feature of this model is the fact that banks offer a variety of services. The simplest example of this is the case of a bank with a large number of branches. Since customers have different preferences over branch location, branches at different locations are offering different services. This means that a bank with many branches is offering a bundle of different services to their customers.

Location is not the only dimension along which banks differ, of course. They will offer different menus of accounts, or concentrate on different types of lending business; they will attract a different mix of retail or wholesale funds; they may diversify into non-bank products such as insurance or mutual funds. Location is a convenient metaphor for these different dimensions.

By exploiting these three features of the model, lock-in effects, limited information, and product diversity, we can reverse the usual presumption that greater concentration leads to more efficient outcomes.

4.1.1 The Model

There is a finite set of locations indexed by $\ell = 1, ..., L$ and at each location there are two banking offices $j = 1, 2$. Time is divided into an infinite number of discrete periods $t = 1, 2, ...$ There is a large number of individuals allocated exogenously to the different locations. Each period, these individuals have a demand for a unit of banking services which provides them with a surplus v. The value of banking services to individuals is distributed according to the distribution function $F(v)$, that is, $F(v)$ is the fraction of the population with valuation less than or equal to v.

At each date consumers are randomly assigned to a new location. They have an equal probability of arriving at any location and we assume that the number of locations is so large that the probability of returning to the same location is negligible and can be ignored. This is an extreme assumption, to be sure, but it serves to eliminate inconvenient and apparently unimportant complications. Each location is also assumed to receive a representative sample of the different types of individuals so, whatever the number of individuals at a given location, the distribution of types is $F(v)$.

For simplicity, banks are assumed to have a zero marginal cost of providing banking services. Profit is thus identical to revenue.

At each date, the market is assumed to clear as follows. First, the individuals who have gathered at a particular location choose which bank to patronize. They do this before they know the price that the bank will charge. Next, the bank sets the price for its product (the interest rate on loans or deposit accounts, or the fees for other bank services). Finally, the consumers make one of three choices: to purchase the current bank's services at the quoted price, to switch to the other bank, whose price by now has been fixed, or to do without the services of a bank. There is a fixed cost $c > 0$ of switching from one banking office to the other.

We consider two limiting cases of bank organization. In the first, which we call *unitary banking*, each bank has a single branch. In other words, each banking office represents

an independent bank. In the second case, which we call *branch banking*, there are only two banks, each owning one branch in each location. That is, all the banking offices are organized into two large networks. Regardless of the form of organization, we use the term banking office to denote the smallest unit of the bank, whether it constitutes the entire bank or a branch of a larger bank network.

Individuals are assumed to observe only what happens at their own bank in each period and, since they move to a different location in each period, they have no knowledge of the previous behavior of the bank they are patronizing in the current period. The banks themselves are assumed to condition their behavior in each location on their experience at the same location. This maintains an informational symmetry between the unitary and branch-banking forms of industrial organization, i.e., unitary banking and branch banking, since in each case only local information is being used to condition the (local) pricing decision.

4.1.2   Unitary Banking

In this form of industrial organization each bank consists of a single office. A bank sets the price of its product in each period to maximize the present value of profits. In the static version of this model, it is well known that the unique equilibrium involves each banking office choosing the monopoly price. More precisely, suppose that there is a unique price p_M such that

$$p_M(1 - F(p_M)) \geq p(1 - F(p)), \forall p.$$

A bank will clearly never want to charge more than p_M. If the two banks at some location happen to charge prices $p \leq p' \leq p_M$, where $p < p_M$, the bank charging p can always raise its price by ϵ without losing any customers, because of the fixed cost of switching. This will clearly increase its profits, so the only equilibrium is for both banking offices to charge p_M.

In a dynamic context things are generally more complicated, because of the possibility of supporting a different equilibrium by means of punishment strategies. Under the maintained assumptions, however, there is no possibility of using such strategies to increase the set of equilibria. Because individuals only observe what happens at their own locations and never return to the same location, nothing that a banking office does in the current period will be observed by individuals who will visit that location in the future. Further, since banks at other locations do not condition their behavior on what happens at this location, there is no possibility of the bank's future customers being indirectly informed of a deviation through another bank's reaction to this bank's current deviation. Our informational assumptions have effectively severed any possible feedback from a current change in price to a future change in demand, so the argument used in the static model continues to apply. We conclude, then, that the unique, subgame perfect equilibrium of the unitary banking game consists of each bank, in each location, charging the monopoly price p_M in every period. Consumers whose valuation v is greater than p_M will purchase banking services; those whose valuation is less than p_M will not. The equilibrium is inefficient for the usual reason: the monopoly price is too high and the monopoly quantity is too low.

4.1.3 Branch Banking

Now suppose that banking offices are formed into two large networks. Each bank has one branch in each of the locations. Although consumers move from location to location they can stay with the same bank if they wish. The possibility of staying with the same bank generates a plethora of other equilibria. We describe one such equilibrium to illustrate the possibilities.

In each period, at each location, half the customers patronize each of the banks. Along the equilibrium path, the banks charge a price $p = \varepsilon > 0$ in every period. If, in any period, one of the banking offices has deviated from the equilibrium strategy, all the customers will leave the bank that last deviated and henceforth patronize branches belonging to the other bank. The banks continue to charge the same price. If no bank deviated, but some of the customers deviated in the past, the banks continue to charge the same price and the customers continue to patronize the two banks equally. The profits (for a single banking office) from deviating last one period are less than or equal to $(p_M - \varepsilon)(1 - F(p_M))$; on the other hand, it loses the profits from these customers in each future period until they all disappear from the game. Customers only last a finite number of periods, but if the number of locations is large, they will be around on average for a very long time. Each period, a fraction ℓ^{-1} of the customers dies and is replaced. So the equilibrium profits lost by a single banking office's deviation are equal to $[\varepsilon/(1-\ell^{-1})(1-\delta)](1-F(\varepsilon))$. For δ sufficiently close to 1 and ℓ sufficiently large, the profits from deviating are less than the profits of the equilibrium strategy. This shows that under branch banking, it is possible to support equilibria which are "more efficient" than the unique equilibrium in the case of unitary banking, where "more efficient" means that the sum of consumers' and producers' surplus is greater.

How should we interpret these results? The lock-in effects in the banking sector may be substantially greater than in most service industries and may be one of its distinguishing features. Small banks with a limited range of services and a limited geographical presence may have a greater incentive to exploit the lock-in effect than a large bank, because the large bank is always competing for the customer's future business, in another product line or another location.

4.1.4 Empirical Evidence

There is some empirical evidence in support of the view of competition presented above.[3] Bordo, Rockoff and Redish (1994) compare the Canadian and US banking systems from 1920 to 1980. During this period Canada had a few branch banks while the US had unit banking in many parts. It is found that the Canadian system outperformed the US system in a number of respects. First, in Canada the interest rates paid on deposits were generally higher and the income received by securityholders was generally slightly higher than in the US. Second, the interest rates charged on loans were generally quite similar in the two countries. Finally, the returns on equity were generally higher in Canada. Taken together this evidence is

---

[3]We are grateful to Stephen Haber for bringing this evidence to our attention.

consistent with the Canadian branch banking system being more competitive than the US unitary banking one.

Carlson and Mitchener (2003) also find evidence that branch banking is more competitive than unitary banking. Using data on national banks from the 1920s and 1930s, they compare states which just have unit banks with ones that have both branch banking and unitary banking. In the former states, many banks that charge high rates of interest on loans and pay low rates on deposits survive. However in states with both types of banks this kind of local monopoly is eliminated. Both branch banks and unit banks price in exactly the same way.

4.2   Spatial Competition and Diversification

We next consider another model of spatial competition in the tradition of Hotelling. It is shown that competition can be consistent with diversification and hence stability. However, the precise way in which this works depends crucially on the particular assumptions made.

Let locations be denoted by $n = 0, \pm 1, \pm 2, ...$ and assume there is a single bank at each location. Identical individuals are uniformly distributed on the real line. Each has one unit of a good that can be invested by the bank in a risky asset with return R_n. For simplicity assume that the returns R_n are i.i.d. with distribution

$$R_n = \left\{ \begin{array}{ll} R & \text{w.pr. } \pi \\ 0 & \text{w.pr. } 1 - \pi. \end{array} \right.$$

In a symmetric equilibrium a bank will draw clients from the interval $[n - 1/2, n + 1/2]$ and offer them a risk sharing contract that promises a payment D if the project is successful and nothing otherwise (the banker is risk neutral but has limited liability). The expected utility of the contract is $\pi U(D) + (1 - \pi)U(0) = \pi U(D)$ if $U(0) = 0$.

The bank chooses D to maximize profits taking as given the expected utility $\bar{u}$ offered by the adjacent banks. Assuming linear transportation costs of one util per unit distance, the marginal agent $m < n$ who goes to bank n is determined by the condition that $\pi U(D) - (n - m) = \bar{u} - (m - n + 1)$ or

$$2(n - m) = \pi U(D) - \bar{u} + 1.$$

The profit per agent is $\pi(R - D)$ so total profits are

$$\pi(R - D)2(n - m) = \pi(R - D)\left(\pi U(D) - \bar{u} + 1\right).$$

The first-order condition is
$$\pi U'(D) = \frac{1}{R - D}.$$

If we allow a bank to occupy several locations, it can pool several independent risks, which allows it to offer better risk sharing to its customers. However, it may face less competition. Whether it does depends on precisely which locations a bank is allowed to

19

occupy. For example, if two banks occupy alternate locations, we get improved risk sharing with no loss of competition. Here there is no trade-off between competition and stability. If a bank occupies a sequence of adjacent locations, it can extract a large amount of surplus from the consumers located in the middle. Here there is a trade-off between competition and stability.

5   Schumpeterian Competition

Technological innovation is one of the major sources of growth in welfare. As Schumpeter (1950) famously pointed out, perfect competition undermines the incentive to innovate (when intellectual property rights are weak) and in that sense imperfect competition may be more "efficient" than perfect competition. Similar ideas apply in the financial sector. If banks innovate they may be able to capture the market and drive other banks out of business. Thus Schumpeterian competition may be associated with financial instability (creative destruction). We start by considering a "winner-takes-all" model of competition based on Allen and Gale (2000c) which has this feature.

5.1   Winner-Takes-All Competition

Firms are set up by teams of managers. There is an exogenously specified set of teams $i = 1, ..., n$ and each team sets up a single firm. The aim of the firm/team is to develop a product that will be produced and sold in the future. The value of the product depends on the amount of capital invested and on the quality of the management team.

There are two dates, $t = 0, 1$. At date 0, the management of firm i chooses an amount of capital $k_i \geq 0$ to invest in the development of a product. Investment in the firm is provided by a large number of risk-neutral, outside investors, who play no active role in the management of the firm. The opportunity cost of capital is denoted by R: for every unit of capital invested at date 0 the investors give up a safe return of R units at date 1.

At date 0 there is uncertainty about the quality of the management team and the decisions they make. However, there is no *asymmetry* of information: managers and investors have the same information about the quality of the management team and, hence, about the future payoffs of the firm. This uncertainty is resolved at date 1 when the quality of the management teams and their products is revealed. Uncertainty about the quality of the teams is represented by the state of nature ω, which is unknown at date 0 and becomes common knowledge at date 1.

The value of the product developed by team i is given by $V_i(k_i, \omega)$. Because there is symmetric information about the quality of the team, insiders and outsiders know the function $V_i(\cdot)$ and the investment k_i and have the same probability distribution $F(\cdot)$ over the states of nature ω. We assume that $V(0, \omega) = 0$ for all ω so capital is essential to the development of a useful product. In general, the more capital that is provided the greater is the probability that the value of the product is high. We assume that once the new product is developed it

can be produced at constant marginal cost and, without essential loss of generality, we set the marginal cost equal to zero.

Managers are assumed to be risk neutral and maximize the expected terminal value of the investors' net wealth. Note that there is no moral hazard problem. The managers work as hard as they can to maximize the expected returns of the investors. They also commit to pay all the profits to the investors.

At date 1 there is assumed to be a continuum of identical consumers whose measure is normalized to one. Each consumer wants to consume at most one unit of a new product.

5.2   Equilibrium

There is a finite number of locations $n = 1, ..., N$ with a bank at each location. There are two dates, $t = 0, 1$. At date 0, each bank n invests $k_n \geq 0$ in the development of a product. The banks' opportunity cost of capital is R. The value of the product developed by bank i is given by $V_i(k_i, \omega)$. There is symmetric information all agents know the function $V_i(\cdot)$ and the investment k_i and have the same continuous probability distribution $F(\cdot)$ over the states of nature ω. We assume that $V(0, \omega) = 0$ for all ω so capital is essential to the development of a useful product. In general, the more capital that is provided the greater is the probability that the value of the product is high. We assume that once the new product is developed it can be produced at constant marginal cost and, without essential loss of generality, we set the marginal cost equal to zero.

At date 1 identical consumers are uniformly distributed on the line interval $[0, N]$. Each consumer wants to consume at most one unit of a new product.

5.3   Equilibrium

At date 0 the firms jointly choose their investment strategies $k = (k_1, ..., k_n)$. At the beginning of date 1, the state of nature ω is realized and the banks observe the quality of the product they have developed

$$V(k, \omega) = (V_1(k_1, \omega), ..., V_n(k_n, \omega)).$$

Then the banks engage in Bertrand competition. Since the qualities are assumed to be continuously distributed (for $k_i > 0$) the probability of ties can be ignored. Then Bertrand competition will lead to an outcome in which the best product captures the entire market and the price charged for this product is equal to the difference between the value of the first and second-best products. The price for every other product is zero. At this price, consumers are indifferent between the first and second-best products, but they will demand only the first-best product in equilibrium (if a positive fraction of consumers were expected to choose the second-best product, the firm with the first-best product would have chosen a slightly lower price to capture the entire market). Formally, for any bank i let

$$V_{-i}(k_{-i}, \omega) = (V_1(k_1, \omega), ..., V_{i-1}(k_{i-1}, \omega), V_{i+1}(k_{i+1}, \omega), ..., V_n(k_n, \omega))$$

21

denote the vector of the qualities of products $j \neq i$; let

$$k_{-i} = (k_1, ..., k_{i-1}, k_{i+1}, ..., k_n)$$

denote the allocation of investment in products $j \neq i$; and let

$$V_{-i}^*(k_{-i}, \omega) = \max_{j \neq i}\{V_j(k_j, \omega)\}$$

denote the highest value in the vector $V_{-i}(k_{-i}, \omega)$. Then, in the second-period equilibrium, the price charged for the i-th product is denoted by $p_i(k, \omega)$ and satisfies

$$p_i(k, \omega) = \max\{V_i(k, \omega) - V_{-i}^*(k_{-i}, \omega), 0\}, \forall i.$$

Since the demand is equal to one for the best product and zero for the rest, the revenue of bank i is also equal to $p_i(k, \omega)$.

At the first date, we look for a Nash equilibrium in the investment levels. The i-th bank chooses k_i to maximize $E[p_i(k_i, k_{-i}, \omega)] - Rk_i$, taking as given the investment levels of the other banks, k_{-i}. So a Nash equilibrium is a vector k^* such that

$$k_i^* \in \arg\max_{k_i \geq 0}\{E[p_i(k_i, k_{-i}^*, \omega)] - Rk_i\},$$

for each i.

5.4   Optimum

Since the cost of production at date 1 is zero, the surplus generated by consuming the i-th product is $V_i(k_i, \omega)$. Surplus is maximized by having all consumers consume the best product, so the total surplus at date 1 is

$$V^*(k, \omega) \equiv \max_{i=1,...,n}\{V_i(k_i, \omega)\}.$$

Assuming that the consumers are also risk neutral and that lump sum transfers are possible, the first-best efficient allocation is found by maximizing net surplus, that is, by solving the planner's problem:

$$\max_{k \geq 0} V^*(k) - R\sum_{i=1}^{n} k_i,$$

where $V^*(k) \equiv E[V^*(k, \omega)]$ is the expected value of $V^*(k, \omega)$.

Define $V_{-i}^*(k_{-i}) \equiv E[V_{-i}^*(k_{-i}, \omega)]$. Then the objective function $V^*(k)$ can be written equivalently as

$$
\begin{aligned}
V^*(k) &= E[\max\{V_i(k_i, \omega) - V_{-i}^*(k_{-i}, \omega), 0\} + V_{-i}^*(k_{-i}, \omega)] \\
&= E[\max\{V_i(k_i, \omega) - V_{-i}^*(k_{-i}, \omega), 0\}] + V_{-i}^*(k_{-i})
\end{aligned}
$$

22

and the planner's problem can be rewritten equivalently as

$$\max_{k \geq 0} E[\max\{V_i(k_i, \omega) - V^*_{-i}(k_{-i}, \omega), 0\}] + V^*_{-i}(k_{-i}) - R \sum_{i=1}^{n} k_i$$

Suppose that k^* is a solution to the planner's problem above. A necessary condition is that k^*_i maximizes

$$E[\max\{V_i(k_i, \omega) - V^*_{-i}(k^*_{-i}, \omega), 0\}] - Rk_i$$
$$= E[p_i(k_i, k^*_{-i}, \omega)] - Rk_i.$$

But this means that k^*_i satisfies the equilibrium condition for the i-th bank's choice of k_i. Hence, we have the following result.

Proposition 6 *If k^* is a solution to the planner's problem, then k^* is a Nash equilibrium of the banks' investment "game".*

In this case the allocation produced by the winner-takes-all competition is efficient. The innovating banks have exactly the right incentives to invest. However, only one bank survives in each period. In other words there is considerable financial instability. In order to prevent this instability the government may wish to restrict competition and give each bank a monopoly in a particular region. We turn next to this restricted competition.

5.5   Restricted Competition

Next assume that the bank located at n has a legal monopoly of the region $[n-1/2, n+1/2]$. Then it maximizes $E[V(k_n, \omega) - Rk_n]$. Clearly, surplus must be lower, both because we are not providing the best product to all consumers but also because we are not using the right investment level.

Will investment be too high or too low? Examples can be constructed to show that the answer is ambiguous. On the one hand losers are protected under the "competitive" arrangement. On the other hand winners suffer because they cannot capture the entire market.

To see this consider the following simple example. There is a continuum of identical consumers with measure one and two banks. Each bank can invest 0 or $0 < I < 1/2$. If a bank invests 0 the value of its good is 0; if it invests I the value of its good is 1. If the market is unified, there are three equilibria, two asymmetric equilibria in which only one bank invests and a mixed strategy equilibrium in which the probability of investment is $\lambda = 1 - I$. Now suppose we divide the market between the two banks, each getting $1/2$ of the market. Then each bank will invest for sure, since $I < 1/2$.

Comparing equilibria, we see that total investment is greater in the divided market. Also, comparing the unique pure strategy equilibrium of the divided market with the mixed strategy of the unified market, we can see that (the probability of) innovation is strictly

higher in the divided market. In the mixed strategy equilibria of the divided market the probability of innovation is the same.

For the case $I > 1/2$ there will be no investment, and hence no innovation, in the segmented market. In this case investment and innovation are unambiguously lower than in winner-takes-all case.

These examples are enough to indicate that even in simple models it is hard to obtain robust comparative static results. As with the other models we have examined, the relationship between competition and stability is complex and nuanced.

6   Contagion

One source of instability in financial systems is the possibility of *contagion*, in which a small shock that initially affects one region or sector or perhaps even a few institutions, spreads from bank to bank throughout the rest of the system, and then affect the entire economy. There are a number of different types of contagion that have been suggested in the literature. The first is contagion through interlinkages between banks and financial institutions (see, e.g., Rochet and Tirole (1996a, b), Freixas and Parigi (1998), Freixas, Parigi and Rochet (2000), and Allen and Gale (2000b) for theoretical analyses and Van Rijckeghem and Weder (2000) for empirical evidence). The second is contagion of currency crises (see, e.g., Masson (1999), Eichengreen, Rose and Wyplosz (1996) and Glick and Rose (1999)). The third is contagion through financial markets (see, e.g., King and Wadwhani (1990), Kyle and Xiong (2001) and Kodres and Pritsker (2002)).

The notion of *financial fragility* is closely related to that of contagion. When a financial system is fragile a small shock can have a big effect. A financial crisis may rage out of control and bring down the entire economic edifice. (see, e.g., Kiyotaki and Moore (1997), Chari and Kehoe (2000), Lagunoff and Schreft (2001) and Allen and Gale (2003b)).

In this section we are interested in the relationship between contagion and financial fragility and competition. Allen and Gale (2000b) develop a model of financial contagion through the interbank market. They assumed perfect competition. In that case a small aggregate shock in liquidity demand in a particular region can lead to systemic risk. Although the shock is small it may cause a bank to go bankrupt and liquidate its assets. This in turn causes other banks which have deposits in it to also go bankrupt and so on. Eventually all banks are forced to liquidate their assets at a considerable loss. Perfect competition in the interbank market plays an important role in this contagion. Since each bank is small, acts as a price taker and assumes its actions have no effect on the equilbrium, no bank has an incentive to provide liquidity to the troubled bank.

Saez and Shi (2002) have argued that if banks are limited in number they may have an incentive to act strategically and provide liquidity to the bank that had the original problem. This will prevent the contagion make the banks providing funds in this way better off. The formal model of this phenomenon that captures the role of imperfect competition is similar to the model in Bagnoli and Lipman (1989). They consider the problem of the provision of public goods through private contributions. There is a critical level of resources required

to provide the public good. Each person becomes pivotal and in this case the public good can be provided. Similarly, here each bank would be pivotal in the provision of liquidity to prevent contagion.

Suppose there are two banks, each holding A units of an illiquid asset and M units of money. One unit of the asset is worth one unit if liquidated today and R units if liquidated tomorrow. Money is storable, so one unit of money today is worth one unit tomorrow. Each bank owes one unit of money to a depositor who withdraws today and the liquidation value of the portfolio to a depositor who withdraws tomorrow. In addition, bank 1 owes bank 2 B units today. Suppose that bank 1 has λ early consumers and $1 - \lambda$ late consumers. If $\lambda + B > M + A - \frac{1-\lambda}{R}$, then bank 1 cannot meet its commitments on its own. If it fails, however, bank 2 can claim only a fraction $\frac{B}{1+B}(M + A) < B$. Bank 2 has several actions it can take. It can forgive part of its debt. It can wait for payment at date 2. It can make a cash transfer to bank 1 at date 1. If R is big enough, this may be better for bank 2 than enforcing its debt now.

The same argument obviously applies to any number of banks, though the coordination problem will become more severe as the numbers increase. If there is a continuum of banks or asymmetric information, it may be impossible to sustain the cooperative equilibrium. Note that even in the case with a finite number of banks there may be a coordination failure: if every bank thinks the others will not contribute anything, it may be optimal not to contribute anything (this may depend on the extensive form, e.g., simultaneous moves rather than sequential moves).

Gradstein, Nitzan and Slutsky (1993) show that Bagnoli and Lipman's result is not very robust to the introduction of uncertainty. It remains to be seen whether a model of contagion and imperfectly competitive banks would also not be robust.

A model of an imperfectly competitive interbank market may therefore be more stable than the case where there is perfect competition. As in the original agency model of Keeley (1990) there is again a trade-off between competition and financial stability.

7　Concluding remarks

In this paper we have considered a variety of different models of competition and financial stability. These include general equilibrium models of financial intermediaries and markets, agency models, models of spatial competition, Schumpeterian competition and contagion. There is a very wide range of possibilities concerning the relationship between competition and financial stability. In some situations there is a trade-off as is conventionally supposed but in others there is not. For example, with general equilibrium and Schumpeterian models efficiency requires the combination of perfect competition and financial instability.

There is a large policy literature based on the conventional view that there is a trade-off between competition and stability. Since competition is generally viewed as being desirable because it leads to allocational efficiency, this perceived trade-off lead to calls for increased regulation of the banking sector to ensure the coexistence of competition and financial stability. The most popular instrument for achieving this end was the imposition of minimum

capital requirements on banks. If the owners of banks were forced to put up significant amounts of capital, they would be unwilling to take risks because they would again stand to loose large amounts of funds. The Basel Agreement of 1988 imposed capital controls on banks so that the incentive to take risks would be reduced and they could compete on equal terms. A large literature has developed analyzing the effect of capital controls. For example, Hellman, Murdock and Stiglitz (2000) showed in the context of a simple model of moral hazard that capital controls were not sufficient. In addition to capital controls, deposit rate controls were also necessary to achieve Pareto efficiency.

Our analysis suggests that the issue of regulation and its effect on competition and financial stability is complex and multi-faceted. Careful consideration of all the factors at work both at a theoretical and empirical level is required for sound policy.

References

Allen, F. and D. Gale. (1994). "Limited Market Participation and Volatility of Asset Prices." American Economic Review 84, 933-955.

Allen, F. and D. Gale. (2000a). Comparing Financial Systems. Cambridge, MA.: MIT Press.

Allen, F. and D. Gale. (2000b). "Financial Contagion." Journal of Political Economy 108, 1-33.

Allen, F. and D. Gale. (2000c). "Corporate Governance and Competition." In Corporate Governance: Theoretical and Empirical Perspectives, edited by X. Vives, pp. 23-94. Cambridge, UK: Cambridge University Press.

Allen, F. and D. Gale. (2003a). "Financial Intermediaries and Markets." Working Paper 00-44-C, Wharton Financial Institutions Center. Forthcoming in Econometrica.

Allen, F. and D. Gale. (2003b). "Financial Fragility, Liquidity and Asset Prices." Working Paper 01-37-B, Wharton Financial Institutions Center.

Bagnoli, M. and B. Lipman. (1989). "Provision of Public Goods: Fully Implementing the Core through Private Contributions." Review of Economic Studies 56, 583-601.

Beck, T., A. Demirguc-Kunt and R. Levine. (2003). "Bank Concentration and Crises." Working paper, World Bank.

Berger, A. and D. Humphrey. (1997). "Efficiency of Financial Institutions: International Survey and Directions for Future Research." European Journal of Operational Research 98, 175-212.

Besanko, D. and A. Thakor. (1992). "Banking Regulation: Allocational Consequences of Relaxing Entry Barriers." Journal of Banking and Finance 16, 909-932.

Bordo, M., H. Rockoff, and A. Redish. (1994). "The U.S. Banking System From a Northern Exposure: Stability versus Efficiency." Journal of Economic History 54, 325-341.

Boyd, J. and G. De Nicolo. (2002). "Bank Risk-Taking and Competition Revisited." Working paper, Carlson School of Management, University of Minnesota.

Carlson, M. and K. Mitchener. (2003). "Branch Banking, Bank Competition and Financial Stability." Working paper, Department of Economics, Santa Clara University.

Canoy, M., M. van Dijk, J. Lemmen, R. de Mooij and J. Weigand. (2001). Competition and Stability in Banking. The Hague, Netherlands: CPB Netherlands Bureau for Economic Policy Analysis.

Carletti, E. and P. Hartmann. (2002). "Competition and Financial Stability: What's Special About Banking?" Forthcoming in Monetary History, Exchange Rates and Financial Markets: Essays in Honour of Charles Goodhart Volume 2, edited by P. Mizen, Cheltenham, UK: Edward Elgar.

Chari, V. and P. Kehoe. (2000). Financial Crises as Herds." Working paper, Federal Reserve Bank of Minneapolis.

Diamond, D. and P. Dybvig. (1983). "Bank Runs, Deposit Insurance, and Liquidity." Journal of Political Economy 91, 401-419.

Diamond, P. (1971). "A Model of Price Adjustment." Journal of Economic Theory 3, 156-168.

Eichengreen, B., A. Rose and C. Wyplocz. (1996). "Contagious Currency Crises: First Tests." Scandinavian Journal of Economics 98, 463-484.

Freixas, X. and B. Parigi (1998). "Contagion and Efficiency in Gross and Net Interbank Payment Systems." Journal of Financial Intermediation 7, 3-31.

Freixas, X., B. Parigi and J. Rochet. (2000). "Systemic Risk, Interbank Relations and Liquidity Provision by the Central Bank." Journal of Money, Credit and Banking 32, 611-38.

Freixas, X. and J. Rochet. (1997). Microeconomics of Banking. Cambridge, MA.: MIT Press.

Glick, R. and A. Rose. (1999). "Contagion and Trade: Why Are Currency Crises Regional?" Chapter 9 in The Asian Financial Crisis: Causes, Contagion and Consequences, edited by P. Agénor, M. Miller, D. Vines and A. Weber, Cambridge, UK: Cambridge University Press.

Gradstein, M., S. Nitzan and S. Slutsky. (1993). "Private Provision of Public Goods under Price Uncertainty." Social Choice and Welfare 10, 371-382.

Hellmann, T., K. Murdock and J. Stiglitz. (2000). "Liberalization, Moral Hazard in Banking, and Prudential Regulation: Are Capital Requirements Enough?" American Economic Review 90, 147-165.

Hoggarth, G. and V. Saporta. (2001). "Costs of Banking System Instability: Some Empirical Evidence." Financial Stability Review, June 2001, 148-165.

Jensen, M. and W. Meckling. (1976). "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure." Journal of Financial Economics 3, 305-60.

Keeley, M. (1990). "Deposit Insurance, Risk and Market Power in Banking." American Economic Review 80, 1183-1200.

King, M. and S. Wadhwani. (1990). "Transmission of Volatility Between Stock Markets." Review of Financial Studies 3, 5-33.

Kyle, A. and W. Xiong. (2001). "Contagion as a Wealth Effect." Journal of Finance 56, 1401-1440.

Kiyotaki, N. and J. Moore. (1997). "Credit Chains." Journal of Political Economy 105, 211-248.

Kodres L. and M. Pritsker. (2002). "A Rational Expectations Model of Financial Contagion." Journal of Finance 57, 768-99.

Lagunoff, R. and S. Schreft. (2001). "A Model of Financial Fragility." Journal of Economic Theory 99, 220-264.

Masson, P. (1999). "Contagion: Monsoonal Effects, Spillovers and Jumps Between Multiple Equilibria," Chapter 8 in The Asian Financial Crisis: Causes, Contagion and Consequences, edited by P. Agénor, M. Miller, D. Vines and A. Weber, Cambridge, UK: Cambridge University Press.

Matutes, C. and X. Vives. (1996). "Competition for Deposits, Fragility and Insurance." Journal of Financial Intermediation 5, 184-216.

Matutes, C. and X. Vives. (2000). "Imperfect Competition, Risk Taking, and Regulation in Banking." European Economic Review 44, 1-34.

Perotti, E. and J. Suarez. (2003). "Last Bank Standing: What do I gain if You Fail?" Forthcoming European Economic Review.

Rochet, J. and J. Tirole. (1996a). "Interbank Lending and Systemic Risk." Journal of Money, Credit and Banking 28, 733-762.

Rochet, J. and J. Tirole. (1996b). "Controlling Risk inPayment Systems." Journal of Money, Credit and Banking 28, 832-862.

Saez, L. and X. Shi. (2002). "Liquidity Pools, Risk Sharing and Financial Contagion." Working Paper 02-03, Financial Intermediation Research Society.

Schumpeter, J. (1950) "Capitalism, Socialism and Democracy (3rd edition)." New York: Harper & Row.

Van Rijckeghem, C. and B. Weder (2000). "Spillovers Through Banking Centers: A Panel Data Analysis," IMF Working Paper WP/00/88, Washington D.C.: International Monetary Fund.

Table 1

Average Cumulative Fiscal Costs of Banking Crises in 24 Crises, 1977-2000

	Number of crises	Non-performing loans (percentage of total loans)	Fiscal costs of banking resolution (percentage of GDP)
All countries	24	22	16
Emerging market countries	17	28	17.5
Developed countries	7	13.5	12
Banking crisis alone	9	18	4.5
Banking and currency crises of which	15	26	23
Emerging market countries	11	30	25
Developed countries	4	18	16
Banking and currency crises with previous fixed exchange rate of which	11	26	27.5
Emerging market countries	8	30	32
Developed countries	3	18	16

Source: Hoggarth and Saporta (2001, p. 150).

Table 2

Output Losses Associated with Banking Crises, 1977-1998

	Number of crises	Average crisis length (years)	Average cumulative output losses (percentage of GDP)
All	43	3.7	16.9
Single banking crises	23	3.3	5.6
Twin banking and currency crises	20	4.2	29.9
Developed countries	13	4.6	23.8
Emerging market countries	30	3.3	13.9

Source: Hoggarth and Saporta (2001, p. 155).