Förord

Finansmarknadsutredningen (Fi 1997:14) tillsattes i oktober 1997 för att utreda den svenska finansiella sektorns internationella konkurrenskraft.

Utredningen har givit ett antal svenska och utländska experter i uppdrag att utarbeta specialstudier till utredningen. Dessa 27 studier har löpande offentliggjorts via utredningens hemsida på internet och redovisas som separata bilagor till betänkandet. Bilagorna har i vissa fall som syfte att ge en sammanfattande bild över forskningsläget på viktiga områden, i andra fall till att djupare analysera vissa centrala frågeställningar. De analyser och slutsatser som läggs fram i dessa bilagor är respektive författares egna och sammanfaller inte nödvändigtvis med utredningens ståndpunkt.

De 27 specialstudierna finns samlade i fyra volymer. Volym A innehåller fem bilagor som alla granskar finanssektorns betydelse för tillväxten. I volym B:s nio bilagor avhandlas de olika funktioner som finanssektorn har i samhällsekonomin. Volym C:s sju bilagor behandlar en mängd övriga frågor som t.ex. skatter, arbetskraft och regleringar. Volym D innehåller de sex bilagor som har skrivits på engelska.

Stockholm i mars 2000

Bengt Dennis
Särskild utredare

Jonas Niemeyer
Utredningssekreterare

Hans Bäckström
Utredningssekreterare

Innehållsförteckning

Financial Systems, Industrial Structure, and Growth

by

Raghuram G. Rajan [*]

and

Luigi Zingales [**]

Supplement number 23 to the Government Inquiry on the International Competitiveness of the Swedish Financial Sector.

September, 1999

[*] Joseph L. Gidwitz Professor of Finance, University of Chicago, Financial Management Area, Graduate School of Business. 1101 East 58th Street, Chicago, IL 60637, USA. E-mail: raghuram.rajan@gsb.uchicago.edu.
[**] Professor of Finance, University of Chicago, Financial Management Area, Graduate School of Business. 1101 East 58th Street, Chicago, IL 60637, USA. E-mail: luigi.zingales@gsb.uchicago.edu.

Table of Contents

Kort sammanfattning (in Swedish)

Hur påverkar utvecklingen av den finansiella sektorn den industriella tillväxten? Vilken effekt har den på industrins sammansättning och företagens storleksfördelning? Hur skiljer sig olika typer av finansiella system när det gäller sårbarhet i krislägen? Denna studie diskuterar och försöker ge svar på denna typ av frågor, baserat på aktuell forskning.

Den första delen av denna studie fokuserar på de teoretiska och empiriska argumenten varför finansiell utveckling underlättar industriell tillväxt. I den andra delen jämförs två typer av finansiella system: det institutionella, relationsbaserade systemet och det marknadsbaserade ”arm’s length”-systemet utifrån vad som förefaller mest gynnsamt för industriell tillväxt. Speciellt fokuseras på vad som ser ut att vara en ny framgångsrik hybrid – riskkapitalfinansiering. I den sista delen ställs frågan hur olika slag av system reagerar vid makroekonomiska störningar och systemrisker.

En viktig slutsats är att utvecklingen av ett lands finansiella marknader och finansinstitut dramatiskt kan öka tillväxten i branscher och företag som behöver en stor del extern, långfristig finansiering. Särskilt intressant är att finansiell utveckling leder till uppkomsten av fler företag inom branscher vars tillgångar till stor del är immateriella som forskning och utveckling. Detta gäller i hög grad för små teknik-baserade företag. Följaktligen underlättar finansiell utveckling struktur-utvecklingen i den industriella sektorn.

Executive Summary

How does the development of the financial sector affect industrial growth? What effect does it have on the composition of industry, and the size distribution of firms? What is the relative importance of financial institutions and financial markets, and does it depend on the stage of economic growth? How do financial systems differ in their vulnerability to crisis? This study attempts to provide an answer to these questions based on the current state of empirical research. This is not, however, meant to be a comprehensive survey. Instead, we hope to present our own point of view, bolstered by the available empirical evidence.

The first section of this study will focus on the theoretical rationale and empirical evidence for why financial development aids industrial growth. In the second section, we compare and contrast two types of financial system: the institution-heavy relationship based system, and the market-intensive arm's length system. We will ask which type of system is more suitable for industrial growth. In particular, we will focus on what seems to be a recent successful hybrid – venture capital financing. In the last section, we will ask how each kind of system responds to macro-economic volatility and systemic risk. We will conclude with some policy conjectures.

Our main conclusions are the development of a country's financial markets and institutions dramatically increases the growth of industries, such as Computers or Pharmaceuticals, that need a lot of external, long-term finance. More interestingly, financial development leads to more firms being born in industries whose assets are largely long term, or in industries where assets are largely intangible ones such as R&D. Thus financial development facilitates the spread of new ideas and leads to the "creative destruction" of the old. By freeing those with ideas from the need for personal wealth to finance them, financial development creates equality of opportunity for everyone in society, something that hitherto has only been a dream.

By increasing the variety and sources of financing, financial development also decreases risk. The sources of financial development, however, are not explored in this study.

Introduction[***]

How does the development of the financial sector affect industrial growth? What effect does it have on the composition of industry, and the size distribution of firms? What is the relative importance of financial institutions and financial markets, and does it depend on the stage of economic growth? How do financial systems differ in their vulnerability to crisis? This study attempts to provide an answer to these questions based on the current state of empirical research. This is not, however, meant to be a comprehensive survey. Instead, we hope to present our own point of view, bolstered by the available empirical evidence.

The first section of this study will focus on the theoretical rationale and empirical evidence for why financial development aids industrial growth. In the second section, we compare and contrast two types of financial system: the institution-heavy relationship based system, and the market-intensive arm's length system. We will ask which type of system is more suitable for industrial growth. In particular, we will focus on what seems to be a recent successful hybrid – venture capital financing. In the last section, we will ask how each kind of system responds to macro-economic volatility and systemic risk. We will conclude with some policy conjectures.

[***] This study was presented at the Symposium on the International Competitiveness of the Swedish Financial Industry organized by the Bank of Sweden Tercentenary Foundation in March 1999. The comments of Erik Berglöf and Jonas Niemeyer are greatly appreciated.

1 Finance development and Growth

There is a long literature debating the impact of finance on growth – dating at least as far back as Schumpeter (1911) – that emphasizes the positive influence of the development of a country's financial sector on the level and the rate of growth of its per capita income. The argument essentially is that the services the financial sector provides – of reallocating capital to the highest value use without substantial risk of loss through moral hazard, adverse selection, or transactions costs – are an important catalyst of economic growth. Early empirical work seemed consistent with this argument. For example, on the basis of data from 35 countries between 1860 and 1963, Goldsmith (1969, p. 48) concludes that ”a rough parallelism can be observed between economic and financial development if periods of several decades are considered” and ”there are even indications in the few countries for which data are available that periods of more rapid economic growth have been accompanied, though not without exception, by an above-average rate of financial development”.

Nevertheless, studies such as these simply suggest correlation. As Goldsmith puts it ”There is no possibility, however, of establishing with confidence the direction of the causal mechanism, i.e., of deciding whether financial factors were responsible for the acceleration of economic development or whether financial development reflected economic growth whose mainsprings must be sought elsewhere.” While Goldsmith was somewhat pessimistic about the possibility of establishing causation, other economists have expressed downright skepticism that financial development is anything but a sideshow to economic development. Robinson (1952, p. 86) is representative of a school, which seems to believe that institutions follow Say’s law of demand creating its own supply when she claims ”where enterprise leads, finance follows”. Others such as Lucas (1988) argue that the importance of financial development for economic growth has been overemphasized.

The importance of financial markets and institutions in a modern economy is undisputed, and fairly well documented elsewhere, see Levine (1997) and Merton and Bodie (1995), for example. What is more hotly debated is whether financial markets and institutions develop automatically to industrial and individual needs, or whether they take an inordinate time, and special circumstances, to build, and cannot be taken

for granted. The evidence suggests the latter. Let us first understand why this might be the case, then examine the evidence.

1.1 Why don't financial markets and institutions appear instantaneously?

Clearly, there are large fixed costs and a physical time to build in setting up financial infrastructure. Furthermore, it may be hard for private enterprise to internalize all the gains to doing so, especially if these gains are spread over a long period of time. If there is no pre-existing financial infrastructure, the economy needs either the fortuitous combination of rich individuals with the will at a time when there is a sudden large demand for financing, or a farsighted government that can pool resources, for the infrastructure to be set up. Chandler (1977) suggests this is, in fact, what happened in the United States. The financial sector, especially investment banks and the corporate bond market, developed to meet the enormous financing needs of railroads in the mid-nineteenth century. The infrastructure was later available to meet the less concentrated needs of industrial firms around the turn of the century.

Even after the set-up costs are incurred, a financial market needs liquidity – the ability for an agent (a seller or borrower) to get full value for her goods. A number of factors determine a market's liquidity but perhaps most important is that a large number of persons think of that particular market as the natural location for their trades. Thus there is a chicken and egg problem with liquidity – people will not trade in a particular market unless they think the market is liquid, but the market will not be liquid unless they trade. This again implies that the continued vitality of a market often requires that rare event that serves to co-ordinate expectations. For example, the United States Government issued Liberty Bonds in large numbers to finance the war effort during the First World War. People who would otherwise not touch a financial security bought these bonds for patriotic reasons. Peach (1941) argues that it was the successful experience these investors had with the bonds that made them willing to invest in arm's length securities issued by corporations, giving corporate securities markets liquidity, and leading to the enormous expansion of these markets in the 1920s.

More generally, economic theory suggests that a market becomes liquid when, somewhat counter-intuitively, there are a large number of uninformed investors willing to grease the wheels of trade. This is because when a market consists only of informed, experienced traders, everyone is trying to second-guess the counterparty's information from

the way they trade, and few trades take place. The reasoning is ”if you are willing to sell me this stock and I know you are smart, I wonder what it is you know that I do not, and I am less willing to buy”. The difficulty of providing uninformed investors the confidence to participate in the market when, by their very nature, they will not collect a great deal of information is, in our view, one of the most important impediments to market liquidity.

Investors obtain such confidence, not from a detailed scrutiny of the financial press but from word-of-mouth about salient events, and articles in the popular press. There is some evidence of this. Initial public offerings of stock that have higher first day returns in the United States attract more analyst following, perhaps because huge initial returns make these “story stocks” that are written about in the popular press attracting investor interest. These stocks also have higher trading volumes, a natural consequence of more uninformed interest.[1] But there is a downside to attracting the uninformed. A broad-based unpleasant experience can equally easily drive them from the market and word of mouth about such a salient event will take a long time to die down. The bottom line is that liquidity is based on past experience, takes a long time to nurture, and could be somewhat fragile for new markets.

Financial contracts also take time to perfect. There is a process of trial and error in which institutions initially, and eventually markets, learn the nuances of a financial contract, how to price it, what covenants to build in, how to withstand legal scrutiny, and how to obtain effective enforcement.[2] The institutions such as exchanges, audit agencies, custodial services, rating agencies, etc., that help the smooth enforcement of arm’s length contracts take time to set up, and more time to perfect their trade in local conditions.

Financial intermediaries also take time to set up and become effective. As Rajan (1998) argues, institutions intermediate where financial contracts are incomplete, either because contingencies cannot be fully foreseen in the case of innovative transactions, or because the contingencies are too complex to describe up front. The institution uses its reputation to bond itself, and its bargaining power to achieve negotiated outcomes that are superior to those that can be obtained through direct transacting in the market place. A good reputation is crucial for an intermediary because this is what convinces customers to trust it when contracts do not fully protect them, but reputations take time to build, see Kreps and Wilson (1982) and Diamond (1989), for

[1] See Rajan and Servaes (1997).
[2] See Rajan (1998).

example.[3] An institution may also be able to bond itself by putting its monetary capital or future rents at stake, but we have already argued that monetary capital may not be available in concentrated form in economies with primitive financial sectors. Moreover, even if available, monetary capital alone may not be particularly effective in providing assurances to customers, see Boot, Greenbaum, and Thakor (1993).

Finally, financial intermediaries take time to build for the same reasons that large industrial firms take time to build, see Rajan and Zingales (1998a). It takes time for employees to learn to work with each other, to build relationships with customers, and to build routines that ensure the smooth functioning of the organization. Moreover, all this learning cannot take place instantaneously, it has to be sequenced, else it could risk overwhelming the command and control capabilities of the organization.

For all these reasons, it is quite possible that financial markets and institutions do not appear on demand, they evolve slowly. Thus the pre-existing state of financial development could have a significant effect on economic growth. Recent evidence seems quite persuasive that this is indeed the case.

1.2 Evidence on Financial Development and Growth

The rekindling of interest on the empirical connection between financial development and growth owes much to King and Levine (1993). They

[3] In every financial innovation, there are outcomes that are not anticipated up front, and have to be worked out amicably between intermediary and client. This is where the intermediary's reputation plays a part. Some intermediaries develop reputations for being un-cooperative, which hurts their business. An incident reported by Euromoney (April 1995, p. 35) is suggestive. In 1987, Texaco sought bankruptcy protection after being ordered to pay $ 10.5 billion in damages to Pennzoil. Even though Texaco was far from insolvent, the filing put it in technical default on a swap contract with Bankers Trust. The details of the swap contract allowed either party to walk away even if they owed money when one party defaulted (this limited-two-way-payment clause is deservedly obsolete). Bankers Trust owed $ 10 million on the swap. The expectation was that it would waive the default given that it was merely technical. Euromoney reports that top BankersTrust management looked at the situation, weighed the bank's relationship with Texaco – not itself known for showing much mercy to its bankers – and took the windfall gain. Texaco has refused to deal with Bankers Trust ever since.

study 80 countries over the period 1960-1989 to see whether the pre-determined component of financial development predicts long run economic growth. They find that beginning-of-decade measures of a country's financial development – such as the ratio of liquid liabilities of the financial system to GDP, the share of domestic credit allocated by banks, or the ratio of domestic credit to private enterprises to GDP – are strongly related to the country's economic growth, capital accumulation, and productivity growth over the subsequent decade. The economic size of the effects is also large. If in 1970, Zaire had increased the share of domestic credit allocated by banks as opposed to the central bank from 26 percent to the mean for developing countries (about 57 percent), Zaire would have grown about 0.9 percent faster each year in the 1970s, and by 1980 per capita GDP would have been 9 percent above its actual level, King and Levine (1993, p. 734).

While the evidence in this paper sheds additional light, it does not lay to rest all doubts about causality. The skeptic could still offer a number of arguments.

First, both financial development and growth could be driven by a common omitted variable such as the propensity of households in the economy to save. Since endogenous savings (in certain macroeconomic models) affects the long run growth rate of the economy, it may not be surprising that growth and initial financial development are correlated. This argument is also hard to refute with simple cross-country regressions. In the absence of a well accepted theory of growth, the list of potential omitted variables that financial sector development might be a proxy for is large, and the explanatory variables to include a matter of conjecture.

Second, there is a potential problem of anticipation. Financial development – typically measured by the level of credit and the size of the stock market – may predict economic growth simply because financial markets anticipate future growth; the stock market capitalizes the present value of growth opportunities, while financial institutions lend more if they think sectors will grow. Thus financial development may simply be a leading indicator rather than a causal factor.[4]

One way to deal with the first problem, that of omitted variables, is to keep effects other than financial development constant. An ingenious paper by Jayaratne and Strahan (1996) does precisely this. Between

[4] Levine and Zervos (1998) attempt to deal with this by using stock market liquidity rather than market capitalization as a measure of financial development. Market expectations could also be built into liquidity, see Shleifer and Vishny (1992), though to a much lesser extent than into market capitalization.

1972 and 1991, a number of states in the United States did away with regulations preventing banks from opening multiple branches within the state. The authors argue that de-regulation was tantamount to a quantum jump in the development of the financial sector within the state, because it allowed scale economies to be realized through bank mergers, old inefficient management to be shunted out, and new management to be given better incentives. They find that annual growth rates increased by 0.51 to 1.19 percentage points a year after de-regulation.

Of course, one could ask again whether financial sector reform leads to economic growth, or vice versa. Did states liberalize anticipating greater need for financing because of economic growth? Jayaratne and Strahan argue that if this was the case, the volume of bank lending should have exploded after de-regulation. It did not! Instead, they argue that de-regulation led to improvements in loan quality, and this led to better growth. It is hard to make the case that state legislatures de-regulated anticipating the improvement in loan quality.

Jayaratne and Strahan essentially follow a difference in differences approach by asking what happens in a state after de-regulation. As a result, concerns about whether the states that de-regulated were somehow special are mitigated. Rajan and Zingales (1998b) follow a similar approach but with an important twist. Instead of examining the effect of changes in financial development on growth as Jayaratne and Strahan do, they examine the differential effect of financial development across industries within a country.

Their rationale is a theoretical one. They argue that the way to make progress on causality is to focus on the details of the theoretical mechanisms through which financial development affects economic growth and empirically document their working. Specifically, theorists argue that financial markets and institutions help a firm overcome problems of moral hazard and adverse selection, thus reducing the firm's cost of raising money from outsiders. So financial development should disproportionately help firms (or industries) typically dependent on external finance for their growth. Such a finding, they argue, could be the ‘smoking gun’ in the debate about causality. There are two virtues to this simple test. First, it looks for evidence of a specific mechanism by which finance affects growth, thus providing a stronger test of causality. Second, it can correct for fixed country (and industry) effects. As a result, the empirical test is less dependent on a specific macroeconomic model of growth.

Rajan and Zingales construct the test as follows. They identify an industry's need for external finance (the difference between investments and cash generated from operations) from data on U.S. firms. Under the

assumption that capital markets in the United States, especially for the large listed firms they analyze, are relatively frictionless, this method allows them to identify an industry's technological demand for external financing. Under the further assumption that such a technological demand carries over to other countries, they examine whether industries that are more dependent on external financing grow relatively faster in countries that, a priori, are more financially developed.

In contrast to prior studies that have used measures of financial development that are based on actual financing – such as the quantity of domestic credit or the size of the stock market – Rajan and Zingales use the accounting standards in that particular country. This has the advantage of measuring the potential of the financial infrastructure rather than its past use.

Their hypothesis would imply that, ceteris paribus, an industry such as Drugs and Pharmaceuticals, which requires a lot of external funding, should develop relatively faster than Tobacco, which requires little external finance, in countries that are more financially developed. Consider, for instance, Malaysia, Korea, and Chile, which are moderate-income fast-growing countries, that differ considerably in their financial development. Consistent with their hypothesis, in Malaysia, which was the most financially developed by their metric, Drugs and Pharmaceuticals grew at a 4 % higher annual real rate over the 1980s than Tobacco (the growth rate for each industry is adjusted for the worldwide growth rate of that industry). In Korea, which was moderately financially developed, Drugs grew at a 3 % higher rate than Tobacco. In Chile, which was in the lowest quartile of financial development, Drugs grew at a 2.5 % *lower* rate than Tobacco. So financial development seems to affect relative growth rates of industries in the way predicted.[5]

[5] Demirguc-Kunt and Maksimovic (1998) also use micro data to develop a test of the influence of financial development on growth. Using firm-level data, they estimate the proportion of firms whose rate of growth exceeds the growth that could have been supported only by internal resources. They then run a cross country regression and find that this proportion is positively related to the stock market turnover and to a measure of law enforcement. While their paper is similar in spirit to Rajan and Zingales (1998b), there are two essential differences. First, their estimate of the internal growth rate of a firm is dependent on the firm's characteristics. While it is potentially more accurate than Rajan and Zingales' measure of external dependence, it is also more endogenous. Second, they focus on between-country differences in the spirit of traditional cross-country regressions, while Rajan and Zingales' focus is on within-country, between-industry differences.

Rajan and Zingales do two additional tests that suggest causality from finance to economic growth. First, they instrument indicators of financial development with measures of the origin's of a country's legal system (whether British, French, German or Scandinavian as classified by La Porta et al., 1998), and with the efficiency of the country's judicial system. These instruments, especially the former, are likely to be pre-determined, and do not reflect development in anticipation of economic growth. The results persist after instrumenting. Second, Rajan and Zingales drop industries that were large at the beginning of the period and are more likely to have been responsible for the country's financial development. They find that industries that are small at the beginning of the period and financially dependent grow faster in countries that are more financially developed than do less dependent industries. This finding in a sample that is unlikely to be responsible for the state of development of the financial markets, suggests that financial development does indeed facilitate growth, and is not simply correlated with it.

Rajan and Zingales' findings indicate that financial markets and institutions may develop to meet the needs of one set of industries, but then facilitate the growth of another younger group of industries, consistent with Chandler's (1977) description of the evolution of financial infrastructure in the United States in the nineteenth century. Goldsmith (1985, p. 2) based on a study of the balance-sheets of twenty countries echoes this when he writes ”The creation of a modern financial superstructure, not in its details but in its essentials, was generally accomplished at a fairly early stage of a country's economic development”.

1.3 Financial Development and Growth: Details

1.3.1 The Kinds of Firms that are Financed

There is now a growing body of work that documents in greater detail the kinds of firms in the modern economy that financial development helps. Carlin and Mayer (1998) present an extremely interesting set of correlations. Using data from 27 industries in 20 OECD countries over the period 1970 to 1995, they find that industries funding a lot of investment with equity tend to grow faster, and do more R&D, in

countries that have better accounting standards. They also find that industries where workers have high skill levels tend to grow faster and do more R&D in countries with better accounting standards.[6] Interestingly, equity financed industries undertake less fixed capital formation in countries with better accounting standards.

In direct contrast to equity financed industries, bank debt financed industries tend to grow more slowly in countries that are more financially developed and also tend to undertake less R&D. Finally, in contrast to what happens in developed countries, bank dependent industries in countries with low GDP grow faster as the banking system develops but are relatively unaffected by improvements in accounting standards.

What do we make of these findings? Once we think of fixed capital as collateral, then interpretation becomes much easier. Typically, equity financed industries tend to have few hard assets, and substantial intangible assets such as growth opportunities, see Myers (1977). In economies with underdeveloped financial markets and institutions, collateral is essential to obtain outside financing. Thus we would expect industries that would optimally use few hard assets if finance were easy to come by, to use more of them in countries with underdeveloped financial systems. But as accounting standards and credit markets develop, equity financed industries would tend to use less fixed capital, as the evidence suggests. In other words, the intangible assets that they typically possess in abundance become easier to finance, and they do not have to distort asset holdings towards fixed capital.

Carlin and Mayer's other findings are also consistent with the above interpretation. Because the intangible assets in equity financed firms become easier to finance as the financial system develops (where development is measured as better accounting standards, or more bank credit), the industry grows faster, and can finance more R&D – one of the biggest sources of intangible assets. Similarly, a highly skilled workforce is an important intangible asset, and the finding that industries with highly skilled workers tend to grow more and do more R&D as financial markets develop corroborates our interpretation. Finally, industries dependent on bank finance in Japan (which is where Carlin and Mayer measure bank dependence) tend to be physical capital

[6] Following Rajan and Zingales (1998), Carlin and Mayer proxy for an industry's dependence on equity finance by the amount of equity that industry uses to finance investment in the United States. They measure the industry's dependence on bank finance by how much that industry uses in Japan, and they measure the industry's use of skilled workers based on how many workers in that industry in Germany are not classified as unskilled.

intensive, smoke-stack industries, see Hoshi, Kashyap, and Scharfstein (1990). When financial markets and institutions are poorly developed, these industries have excess collateral and can invest in intangible activities such as R&D that others would undertake if financial constraints were not important. As finance develops, these industries lose their comparative advantage, and tend to grow less, and do less R&D.

In sum, it appears that industries can raise finance more easily as the financial system develops because physical collateral becomes less important, while intangible assets and future cash flows can be financed.

1.3.2 The Effects of Development of Firm Size and Birth

Industry growth can be decomposed into the growth in the number of firms and the growth in the average size of firms. As Rajan and Zingales (1998b) find, 2/3rd of industrial growth over the 1980s in their sample of 43 countries comes from the growth in the average size of establishments, 1/3rd from the growth in the number of establishments. But in their regressions, they find financial development has almost twice the economic effect on the latter than on the former. So the effect of financial development on both the size and the birth of firms seems important.

In a study of 15 OECD countries, Kumar, Rajan and Zingales (1999) find that industries dependent on external finance tend to have smaller firms on average, as do countries with low levels of financial development. Moreover, financially dependent industries tend to have larger firms in countries with better-developed financial markets. Therefore, even though large firms typically finance more of their growth from internal funds, the lack of development of financial markets does constrain their growth. This may be because even though large firms use external funds to a much smaller extent, at the margin these funds are important for growth. It may also be that large firms need to continuously re-assure existing claimants about additional growth, and only a developed financial system contains adequate tools of corporate governance to enable large firms to provide such assurances.

By constraining firms from growing to their full potential, the lack of financial development could have serious effects. Firms in financially underdeveloped countries could be inefficiently small, or left unable to undertake the necessary takeovers that would allow them to compete in a increasingly border-less world.

The absence of financial development could also prevent the emergence of new ideas. When firms start out, many of their assets are intangible ideas and project opportunities. These ideas typically do not respect the status quo, and often make possible Schumpeterian ”waves of creative destruction”. By allowing intangible assets to be financed, financial development facilitates innovation and change.

We have more evidence in favor of this hypothesis. Kukies (in progress) finds that the growth in the number of establishments (though not the growth in average firm size) is relatively higher for industries that do a lot of R&D, and for industries that have a higher portion of long term assets, in countries where financial markets are more developed. This indicates that it is very hard for emergent firms to finance intangible or illiquid assets without financial market development.

There are also some puzzling pieces of evidence. Some developed countries have both good accounting standards (which, across countries, is strongly positively correlated with stock market capitalization) and a high proportion of bank lending, but in general these countries are exceptions. In fact, Carlin and Mayer find the raw correlation between the two measures of financial development to be negative in their sample. Why is it that bank credit tapers off as stock markets develop?

That economies seem to emphasize either institutions or markets has led economists to classify financial systems as relationship based (or bank dominated) and arm’s length (or market based). What distinguishes these two types of systems? Moreover, do they have different effects when countries are at different stages of growth? Could this explain why bank dependent industries grow faster in bank dominated economies only when the economies are relatively under-developed? To answer these questions, we have to ask how relationship based systems work and how they are different from arm's length systems. This is the subject of the next section.

2 Relationship vs Arm's length systems

Let us begin with a sketch of the salient features of these two kinds of systems. Like all sketches, this one has elements of caricature, but this is the price we have to pay to avoid being distracted by the details.

A financial system has two primary goals: to channel resources to their most productive uses and to ensure that an adequate portion of the return flows to the financier. The latter is, of course, crucial to the former. Without the prospect of an adequate return, funds will not be made available for investment.

Even when it is clear that the social returns from a project are high, it is not necessary that the financier will be able to appropriate an adequate return. The borrower may take actions that improve its private return at the expense of the financier's return. This is the problem of moral hazard. Alternatively, the financier may not be able to distinguish between honest entrepreneurs and knaves, and this problem of adverse selection may again prevent the financier from recovering the amount invested. What is really needed is that the financier have both information, as well as the ability to act on that information.

Relationship based systems ensure a return to the financier by granting her some power over the firm being financed. The sources of power can be varied. It may be purely financial, stemming from the financier's monopoly control over supply of funds to the firm because of its firm-specific information. It may be legal, stemming from the financier's majority ownership of the firm. Or it may come from the product market because the financier is, or controls, a major supplier or customer. Regardless of the source of power, the financier attempts to secure her return on investment by exercising this power. Also, as with every system based on monopoly, barriers to entry are required. For example, financial monopolies could be created by limiting the number of institutions that are chartered, and by limiting the amount of information disclosed in the system so that outside supply of finance is thwarted by lack of transparency.

Contrast this with the arm's-length, Anglo-Saxon, market based system, where the financier is protected by explicit contracts and transparency. In such systems, contracts and associated prices determine the transactions that are undertaken. As a result, institutional

relationships matter less and the market becomes a more important medium for directing/governing the terms of transactions.

An important distinction between these two systems is their different degree of reliance on legal enforcement. Relationship based systems can survive in environments where laws are poorly drafted and contracts not enforced. The relationship is largely self-governing; parties intent on maintaining their ”reputations” honor the spirit of agreements (often in the absence of any written contract) in order to ensure a steady flow of future business within the same network of firms. By contrast, the prompt and unbiased enforcement of contracts by courts is a pre-condition for the viability of a market-based system. Moreover, since contracts are typically hard to write with the wealth of detail necessary to fully govern transactions, it is important that the law offer a helping hand. We discuss in Rajan and Zingales (2000) how the law, especially a common law tradition, can help.[7]

Another distinction between the two systems is the relative importance of transparency. Market-based systems require transparency as a guarantee of protection. In the words of Franklin Roosevelt, ”Sunlight is said to be the best of disinfectants; electric light the most efficient policemen.”[8] By contrast, relationship based systems are designed to preserve opacity, which has the effect of protecting the relationships from the threat of competition. This probably explains the negative correlation between accounting standards and the size of the banking sector that we noted earlier.

2.1 An Example: Credit

Before going further, let us consider the example of a transaction – the extension of credit – in each of the two systems. In a relationship based system, a bank will have close ties with a potential borrowing firm, perhaps because of frequent past contacts or because of ownership links. In assessing the borrowing needs of the firm and its ability to pay interest and principal, the bank will consider not only the firm's current debt-servicing capability, but also its long-term ability to repay, and the various non-contractual levers the bank can push to extract repayment.[9]

[7] See La Porta, et al. (1997).

[8] In Joel Seligman, *The Transformation of Wall Street,* Northeastern University Press, Boston 1995, p. 42.

[9] For example, the bank may refuse to extend a blanket guarantee to the firm’s other creditors, refuse to provide new financing or even take a piece of it, etc.

The interest rate charged will be repeatedly negotiated over time, and may not have a direct relationship to the intrinsic risk of the project.

In an arm's-length system, by contrast, the firm will be able to tap a wider circle of potential lenders because there will be more widespread financial information about it. The loan will be contracted for a specific period, and the interest rate will be a competitive one that will compensate the lender for time and the risk of that particular loan.

Limitations on competition in a relationship system do not just give the financier power, but also strengthen his incentive to co-operate with the borrower. Studies of Japanese keiretsus show that the main banks went out of their way to help financially distressed borrowers. For example, Sumitomo Bank not only effectively guaranteed Mazda's debts when it got into trouble after the first oil shock, but also orchestrated a rescue, in part by exhorting employees within its keiretsu to buy Mazda cars.[10] Sumitomo's incentive to help would have been considerably weaker if Mazda had had the option of giving the lion's share of its business, once it emerged from distress, to some other bank. As this example suggests, the effective limitations on outside competition imposed by the keiretsu system enable lenders to ”internalize” a greater share of the benefits accruing to the borrowers than is possible in an arm's-length, competitive banking environment.

The absence of competition and disclosure in a relationship based system imply that there are really no price signals to guide decisions. Unlike an arm's-length system, where a number of competitive lenders can give a borrowing firm independent assessments of the costs of undertaking a project, the cost a borrower faces in the relationship based system is simply what the relationship lender and the borrower negotiate. Since there can be substantial value created in the relationship, and the negotiation and allocation of this surplus is a function of each party's power, the effective cost of financing can deviate substantially from the true risk-adjusted cost.

2.2 Do relationship based systems always lead to worse investment decisions?

But is this necessarily a bad thing? Are lending and investment decisions always inefficient if the cost of funds differs from their true cost? Are there no redeeming features of a relationship based system? The answer to all these questions is no. In the real world with all its ”imperfections,”

[10] See Hoshi, et al. (1990a).

an imperfect cost of funds can sometimes produce the right investment decisions.

For instance, consider our previous example of a firm in distress. Taking into consideration all the value that the firm adds to society – to workers, customers, and local governments as well as shareholders – the company may be worth saving. But, *in the short run*, the true cost of funding may far exceed what the firm can pay without creating further investment distortions.[11] And in the competitive arm's-length system, a lender may not be able to recoup or ”internalize” enough of the firm's value in the long run to be able to offer it subsidized financing in the short run. So the firm is much less likely to be bailed out in the competitive, arm's-length system. By contrast, a lender in a relationship based system, confident in the strength of the relationship (and the protection it affords from competition), can offer a below-market rate in the short run and then recoup its losses with an above-market rate over the long run when the firm is healthy and can afford high payouts. In sum, relationship banks can be viewed as using their monopoly power to charge above-market rates in normal circumstances in return for an implicit agreement to provide below-market financing when their borrowers get into trouble.

Petersen and Rajan (1995) provide evidence of the existence of such relationship lending practices even in the U.S. In examining bank loans to small businesses in different banking ”markets” throughout the U.S.,[12] their study finds that in ”concentrated” markets (those where most of the lending is done by a handful of banks) – which are likely to be more relationship-oriented for the reasons discussed earlier – more credit is available to young firms than in more competitive banking markets. To the extent young firms are more credit rationed, as many observers have suggested, the evidence suggests that the relationship based system does a better job of ensuring that value-adding projects get funded.

The study also finds that the interest rates charged to younger firms are, on average, lower in concentrated markets than in competitive markets, with the effect reversing for older firms. This suggests that banks in concentrated markets can offer more credit on economic terms because their relationships allow ”intertemporal cross-subsidies” – that is, below-market rates for younger firms that are compensated for by above-market rates for more mature firms that have a higher ability to

[11] For example, too high an interest rate could lead the firm to take riskier, negative NPV, projects, see Jensen and Meckling (1976).

[12] The idea of distinct banking markets makes sense in this case because small firms rarely do business with a bank outside their local banking market; the median borrower in the above cited study is only two miles from its bank.

pay. Such subsidies, as suggested earlier, would not be possible in more competitive markets.

Clearly, it is this kind of ability to ”internalize joint surplus” – that is, to trade off short-run losses for longer-run gains – that led so many observers, including many economists, to defend the efficiency of relationship based systems. But it is easy to see the problems that can arise in such systems.

2.3 The distortions in relationship based systems

2.3.1 Poor price signals

Perhaps most important, the relationship based system does not pay much attention to market or price signals. And this indifference to price signals becomes self-fulfilling. If investment decisions are not driven by prices, then prices become less effective in providing economic directions because they reflect less information.

This is not to say that the arm’s length system is perfect in the allocation of resources. Because outsiders have little power, management can indulge itself far more in empire-building without triggering an intervention by outsiders. This problem has been labeled the “agency costs of free cash flows” by Michael Jensen. The arm’s length system, however, can use takeovers to rectify this problem when it gets excessive.[13] By contrast, the problem of mis-allocation of resources due to the lack of price signals in the relationship system is more severe, because it lacks a self-activating mechanism to correct it. In fact, even if price signals were accurate, the power structures in the relationship based system may not allow movement in a direction indicated by the prices.

Evidence of this unwillingness to respond to market signals was provided by a 1991 study by Hoshi, Kashyap, and Scharfstein.[14] The study looked at a sample of Japanese firms in the late 1970s to mid 1980s that had close ties to banks and compared their investment

[13] If anything, managerial empire-building is less severe in a relationship based system, precisely because financiers have the power to intervene extensively and absorb free cash flows from successful firms.
[14] Hoshi, et al (1991).

behavior with a sample that had no such ties. The investments of firms that had no bank tie were very sensitive to the cash flow the firms generated from operations; when operating cash flows decreased sharply, so did investment spending – and vice versa. By contrast, the investments of firms with strong ties to the banks were significantly less sensitive to the firms' operating cash flow.

As suggested earlier, one possible interpretation of these findings is that banking relationships make it easier for firms to obtain external funding for value-adding investments, thus making them less dependent on their own cash flows. But recent events in Japan suggest a different explanation. More often than not, the companies' continuous access to bank funding on favorable terms allowed them to ignore the signal sent by their poor cash flows, and to continue investing. By continuing to invest in these circumstances, such firms may well have been destroying long-term value rather than increasing or preserving it. Even if the banks were failing to provide the managers of these firms with the right signals, it appears that the stock market was attempting to do so. For, as the study also reported, the firms with banking relationships in their sample had lower ”Tobin's q” (or market-to-replacement cost) ratios than firms without bank ties (consistent with our earlier conjecture that bank dependent firms are asset intensive, low growth firms). And, to the extent Tobin's q is a reliable proxy for a firm's investment opportunities, the stock market was expressing skepticism about the likely payoff from such investments.

Weinstein and Yafeh (1998) suggest that such market skepticism was warranted. For while Japanese firms with close bank ties may have had greater access to funds when their operating cash flows declined, such access did not enable them to achieve higher profits or growth rates than their peers.

Peek and Rosengren (1998) provide additional evidence that relationships can distort the allocation of funds. In the early 1990s, Japanese banks increased their lending to the U.S. commercial real estate market. At their peak in 1992, the U.S. subsidiaries of Japanese banks accounted for one fifth of all commercial real estate loans held in the U.S. banking sector. Then, in response to a severe decline in real estate prices in Japan, the Japanese banks cut back their lending in the U.S. even as U.S. prices were rising (and lending by non-Japanese banks increasing) while at the same time expanding their lending in the domestic Japanese market where prices were plummeting. Thus, rather than cutting their losses in Japan – or at least not abandoning their profitable opportunities in the U.S. – Japanese banks poured more money into their unprofitable Japanese relationships.

In sum, one downside of a relationship based system is that price signals are obscured. The consequence could be a widespread and costly misallocation of resources.

By contrast, there is a virtuous circle at work in market-based economies: In the process of relying on prices for guidance, the arm's-length transactions that predominate in these economies also have the beneficial effect of making prices more informative. Thus, the more transactions that come into the market, the more likely decisions made on the basis of price are likely to be the right ones.

2.3.2 Market Power

Another consequence of prices being obscured in a relationship based system is that the financier's information is largely private especially when the projects being financed consist of intangible assets such as intellectual property. Since the creation of intellectual assets requires substantial endeavor on the part of management, the rent sharing entailed in a relationship system can depress management effort, making innovation, especially of the entrepreneurial kind in high technology industries, a rare commodity.

2.3.3 Illiquidity

Because information is so concentrated in a relationship based system, financial assets become very illiquid. Since the relationship is specific to the intermediary and borrower, the intermediary becomes indispensable to collecting on loans. As a result, the intermediary can extract sizeable rents from investors also, see Diamond and Rajan (1998), because they need the intermediary's skills to realize the value of the intermediary's loan portfolio. Anticipating these rents, investors will charge the intermediary a high cost of capital. Moreover, the financial asset will be illiquid in the sense that it will trade in the market for true value (the amount that can be collected from borrowers) less the rent accruing to the intermediary because of its indispensable relationship specific skills. The illiquidity of financial assets makes it very costly for the financial system if the intermediary gets into trouble since more liquid outsiders cannot take over its assets easily.

2.4 The kinds of assets financed by each system

All this then suggests why pure relationship based systems tend to have a comparative advantage in financing physical asset intensive industries rather than high technology R&D based industries. For one, physical capital intensive industries are typically more traditional and well understood. As a result, the absence of market signals about their profitability is less likely to be a problem in making investment decisions. Second, because they are well understood, it is unlikely that a large amount of rents will accrue to the financing intermediary. Moreover, the borrower has the collateral to entice fresh lenders if the existing one proves demanding. Finally, since loans are well collateralized by physical assets, they are liquid, so the concentration of information is less important a barrier for the intermediary to finance its assets.

Conversely, arm's length systems will have a comparative advantage financing industries with intangible assets, hence Carlin and Mayer's finding that equity- and skill-based industries tend to do more R&D in economies with better developed accounting standards.

An intriguing recent study fortifies our view that relationship based systems are more capable of financing projects where the ratio of tangible to intangible assets is large. Houston and James (1995) study the financing arrangements of 250 public firms in the United States. They find that firms with relationships to single banks tend to use less bank debt in proportion to total debt as their market to book ratio (a measure of the ratio of intangible to tangible assets) increases. By contrast, when firms have relationships with multiple banks, the ones with higher market to book ratios tend to use more bank debt in proportion to total debt. This suggests that firms tend to avoid becoming dependent on a single bank when they have high market to book ratios, perhaps because they fear the bank may have too much power to extract rents and direct strategies, or because the bank itself will find the asset too illiquid.

By contrast, when the firm has multiple banking relationships (effectively an arm's length system), no single bank has too much power. Since public markets can obtain information about the firm from multiple sources, the cost of borrowing more from banks to fund is small relative to the insurance and advice provided by relationships. Thus the revealed preference of firms, in an environment where they can choose the relationship structure that benefits them most, gives us a sense of when a relationship is onerous and when it is not.

This study also suggests the answer to an apparent contradiction. If relationships are inimical to the financing of skill based and high technology industries, then why is it that venture capital, which seems the quintessential form of relationship financing, is so important in their financing even in the most advanced arm's length systems. We believe that the conditions under which venture capital arises ensure that the worst problems of the relationship system are tempered by the close proximity of the arm's length system. The error leading to the apparent contradiction is to view venture capital as purely a form of relationship financing. In fact, it seems an ideal bridge between relationship and arm's length financing, combining the best of both worlds.

2.5 Venture Capital: A bridge between two worlds

Relationship based finance, as we have seen, has the virtue that the financier has substantial control over the financed, and is able to guide the borrower's moves, as well as thwart any malfeasance. How can the venture capitalist bring these virtues to financing high technology without inflicting the costs associated with relationship financing?

In our view, the role of the venture capitalist is to reduce the illiquidity of the financed firm – which is the source of many of the ills of the relationship system – and the existence of a vibrant arm's length market is crucial for him to perform his role. His constant endeavor after financing the firm initially is to prepare to exit. He does this by making the firm's management and control processes standardized and bureaucratic, transparent, and easy for arm's length investors to take over. The venture capitalist also moves the firm from being an organization dependent on the founder to an organization capable of being run by professional managers. This again makes the firm easier for arm's length investors to control since managers are easier to replace than founders.[15]

Why does the venture capitalist not simply hold on to his stake in the firm and extract rents? The answer is that he obtains much greater returns from taking the firm public. The growth opportunities of the typical successful venture capital financed firm are so high that the venture capitalist would not be able to finance them in entirety without severely rationing his other ventures. As a result, unless the venture

[15] For a theoretical analysis of the control role of venture capital, see Berglöf (1994), for an empirical analysis, see Gompers (1995).

capitalist makes the firm easy for outsiders to finance – by establishing standardized management procedures and disclosure which make it easy for outsiders to understand, and exert control over, the firm – many of the growth opportunities the firm has access to will lapse. By contrast, if the venture capitalist does prepare the firm for outside control, and takes the firm public in an Initial Public Offering, the Net Present Value of future growth opportunities will be embedded in the stock price, and the venture capitalist will be amply rewarded for his endeavor through his stock holding. The incentives for the venture capitalist are clear, and it is obvious that the presence of an efficient and liquid stock market that can anticipate the firm's future opportunities is crucial.[16]

Moreover, venture capital partnerships are structured with a limited life, which again serves as a commitment to let go of the firm. The need to exit via the market ensures that prices eventually do matter and discipline the investments that take place. Thus the incentives for innovation, the liquidity, and the price discipline provided by the market are combined through the venture capitalist with many of the benefits of control and long horizons provided by the relationship based system.[17]

A number of countries have tried to seed the process of venture capital, with the hope of emulating the successes of Silicon Valley. Often, this takes the form of making available a fund targeted at high-tech industries, presumably based on the idea that the government can bear risks that private investors may be reluctant to bear. The discussion above points to three other barriers to the emergence of venture capital, all of which may be more important than risk aversion in explaining why private capital does not come forth. First, there should exist industries with a sufficiently high potential growth rate to make venture capital effective – high tech, by itself, is not enough. Second, venture capitalists should possess the rare combination of financial expertise in discerning viable projects from non-viable projects, as well as the managerial expertise in making emerging firms "bureaucratized" and capable of being controlled without killing their innovative spirit. Third, an outside

[16] It may not be necessary for the market to be domestic. If domestic accounting practices and corporate governance are adequate, it may be possible for a foreign market to serve as the liquid exit market for venture capitalists. A case in point is Israeli high tech firms that have been choosing the NASDAQ for their Initial Public Offerings. However, Israel may be a special case.

[17] It is interesting that German banks were performing a similar role to venture capital around the turn of the century, and German public markets were vibrant, Calomiris (1994). It would be useful to understand what changed.

equity market should exist which is capable of reflecting the venture capitalist's contribution in the firm stock price when the firm is taken public (see Black and Gilson (1998) for an early exposition on the market being an essential exit route for venture capitalists). Simply putting traditional bankers in charge of a pool of funds targeted at high tech firms without giving them hands on experience in an existing venture market, and without creating a vibrant equity market for new issues, will most likely doom official venture capital efforts to failure.

In sum, relationship based systems can work very well in the early stages of industrialization where the industries to be financed are physical asset intensive, where the legal system is ineffective, and where skill based or idea based industries are of limited import. But as economies develop and focus more on knowledge intensive industries as engines for growth, a hybrid is perhaps more effective. There is then the need to improve transparency, judicial efficiency, and mechanisms for speedy resolution of financial distress so that arm's length markets can function effectively and aid the process of economic growth.

3 Risk

In the current economic situation where America is ascendant while Japan is mired in bad loans, it is tempting to conclude that the arm's length system dominates other forms. We believe that is a misreading of the evidence. The United States does not have a purely arm's length system, and even so, to conclude that the system is optimal is probably as wrong as the conclusion in the late 1980s that the Japanese relationship based system was worthy of emulation in its entirety. Nevertheless, if there is one thing the arm's length can do better than the relationship system, it is to bear and manage macro-economic risk. Let us understand why.

3.1 Why a relationship based system is more risky

We have already argued that assets financed by relationship based systems tend to be illiquid since there is little transparency and disclosure. Diamond and Rajan (1998) argue that intermediaries can finance such assets at low cost only by issuing a high proportion of demandable claims. Intuitively, intermediaries in relationship based systems finance assets that only they understand. The only reason that they do not absorb a massive amount of rents as a result of their monopoly position is because they credibly commit to pay out collections to depositors. This requires them to issue hard claims – the hardest being demandable claims subject to runs. Thus in the natural course of financing illiquid relationship assets, financial intermediaries have to take on financial risk.

Risk can be mitigated if the intermediation system is well capitalized because capital acts as a buffer. Given the low levels of private capital in emerging economies, historically, the government has created capital for intermediaries by keeping the rates intermediaries pay investors low. This has become infeasible as deregulation and competition has given investors more choice. Consequently, the task of averting the collapse of the system of intermediation in the face of severe macroeconomic volatility has shifted directly to the government. Governments have had to absorb risk by promising the intermediation system capital, implicitly

or explicitly, in case the system is in danger. But the promise of such contingent capital carries with it the risk that intermediaries will collectively attempt to game the system through moral hazard.

In other words, illiquid assets can only be financed by financially fragile intermediaries, who then impose risk on the system. To reduce risk, the government has to promise intermediaries contingent capital, which in turn causes them to bet on the same risks such as real estate or emerging market lending, knowing fully well that they will be rescued if only they sink together.

Moreover, once a relationship based system suffers a severe shock that the government is not able to counter, the flow of credit can collapse quickly. This is because, first, there is a lot of specific knowledge and trust embedded in relationships that cannot be transferred to wealthy unaffected outsiders. The illiquidity of the relationships prevents a quick takeover by, say, foreigners. Second, since property rights are not well established in relationships, it becomes hard to separate healthy unaffected parties from the walking dead. The inefficiency of the judicial system does not help. As a result, the relationship based system tends to share the consequences of an adverse shock somewhat indiscriminately. So, it is no wonder that outside capital does not flow in until the system essentially sorts itself out.

Contrast this with the arm's length system where the accent is on providing small investors the confidence to invest directly in firms. Clearly, such a system is better able to withstand shock, first, because the healthy can be distinguished from the terminally ill after a shock and can be dealt with differently and, second, because unaffected outsiders have the ability to invest and revive the system, as they obtain confidence from the very same channels that inspire confidence in small investors.[18]

3.2 How to reduce risk in a relationship based system

We have argued that elements of both relationships and arm's length financing are needed in the modern economy. Moreover, a combination of both may serve to reduce the risk of a financial meltdown. But what if

[18] This is not to say that markets do not seize up on particular firms. In fact, many banks advertise their relationship business as a port of safe haven for firms affected by a market storm. Nevertheless, we think it highly unusual for all world markets to be irrationally down on an entire country.

the economy is more primitive so that arm's length financing is not possible?

As Levine (1999) shows, the growth of the intermediary sector is correlated with an improvement in creditor rights and contract enforcement. Thus the reliance on relationships, even within the intermediated sector, can diminish as contracting improve. But what if a country urgently needs to reduce risk? Perhaps the quickest, but politically most difficult, way of reducing the aggregate risk of the intermediated sector is to allow massive entry of foreign intermediaries. Using bank level data across a broad cross-section of countries, Demirguc-Kunt, Levine, and Min (1998) show that the extent of foreign bank participation in a domestic economy (as measured by foreign bank assets to total domestic banking assets or the number of foreign banks to total banks) is negatively related to the probability of a financial crisis in that country.

What could explain these findings? Clearly, foreign banks may influence local supervisory or risk management practices, as also local bank efficiency. But they could have other spillover effects. Specifically, foreign banks, because of their vast outside resources, will survive a local crisis. Not only does this give them little incentive to herd on the same risks as local banks, it also gives the government confidence that a financial crisis will not result in a total meltdown. As a result, the government will be less eager to bail out failed local banks. Moreover, the foreign banks offer an avenue for foreign investors to invest in the local economy despite the absence of safeguards necessary for arm's length investing, thus shortening the duration of a purely financial crisis. But taken together, perhaps the most salutary effect is on domestic bank incentives. Knowing that the government will be more reluctant to bail them out, they will be more careful about herding on certain systematic risks. Of course, foreign banks come with their problems including the fact that they are impervious to domestic political compulsions (often a blessing) and they are hard for domestic central banks to control (ditto).

We do not believe that emerging markets should unquestioningly open their doors to all forms of foreign capital. In the absence of well developed markets or foreign intermediaries, foreign capital inflows other than foreign direct investment have to be intermediated into the domestic economy by the domestic banking system. Since foreign capital demands substantial safeguards such as implicit or explicit seniority, and a short maturity, in return for putting money in a relatively opaque system, this simply adds another dimension of risk to the domestic banking system. If inflows are large, there is the risk of a rational large scale search for the exits, if the system is shocked (for example, by

devaluation). Rajan and Zingales (1998c) and Diamond and Rajan (1999) explain why such inflows could have contributed to the Asian financial crisis. Thus in opening up, countries should distinguish between foreign intermediaries that open domestic branches, develop local expertise, are well tied into the system, and thus serve to stabilize it, and foreign portfolio inflows that rely on domestic intermediaries and could overload the system.

4 Conclusion: And the winner is…

We started this study by documenting that there does seem to be a causal relationship between financial development and economic growth. Furthermore, financial development seems to particularly help the financing of firms that typically do not get institutional credit because they lack physical collateral. This lead us to a comparative analysis of relationship based banking systems and arm's length market based systems. We concluded that for the kinds of industries that are now engines of world growth, a hybrid is probably best, though not all the best properties of each system survive in the hybrid. Finally, we argue that the developed legal infrastructure necessary to sustain a large banking system or arm's length markets also tends to diminish risk.

From a policy perspective, it would appear that a country intent on economic development should fix its financial plumbing – specifically its accounting and disclosure system and its legal and bankruptcy codes. In a companion study, we will argue that political forces, more than anything else, prevent the emergence of financial systems that are appropriate for economic activity. On the one hand, this is good news because it suggests that wholesale systemic change – in fundamental legal infrastructure, for example – is not needed to achieve financial development. On the other, this is bad news because political forces are hard to overcome. We believe, however, that the natural increase in competition between firms in different nations will force political forces to eventually give way.

Finally, it is tempting to anoint a specific country as having the best financial system. We want to refrain from this. One reason is that we have so much to learn about how financial systems work, even though we have come a long way in recent years. But perhaps a more important reason is that even a cursory study of a sample of the richest countries in the world will reveal a variety of financial systems. Clearly, there are many paths to gold. While there are signs that indicate convergence to a hybrid (the United States using venture capital and Germany opening the New Market), these are far from conclusive. Moreover, many of the stylized facts about the effects of particular systems – such as the "fact" that bank oriented economies tend to have firms with more debt – are illusory when examined more closely, see Rajan and Zingales (1995). Therefore, we will not go beyond the facts and the interpretations

detailed above, and like all good academics, leave further speculation to future research.

References

Berglöf, E., (1994), ”A control theory of venture capital finance”, *Journal of Law, Economics and Organization*, 10, 247-267.

Black, B. and R. Gilson, (1998), ”Venture Capital and the Structure of Capital Markets: Banks vs Stock Markets”, *Journal of Financial Economics* 47, No. 3, 243-278.

Boot, A., S. Greenbaum, and A. Thakor, (1993), ”Reputation and Discretion in Financial Contracting”, *American Economic Review* 83, 1165-1183.

Calomiris, C., (1994), ”The Costs of Rejecting Universal Banking: American Finance in the German Mirror, 1870-1914”, mimeo, University of Illinois, Urbana Champaign.

Carlin, W., and C. Mayer, (1998), ”Finance, Investment and growth”, mimeo, University College, London.

Chandler, A., (1977), ”The Visible Hand: The Managerial Revolution in American Business”, Harvard University Press: Cambridge MA.

Demirguc-Kunt, A., R. Levine, and H. Min, (1998), ”Opening to Foreign Banks: Issues of Stability, Efficiency, and Growth”, mimeo, World Bank.

Demirguc-Kunt, A. and V. Maksimovic, (1998), ”Law, Finance, and Firm Growth”, *Journal of Finance*, 53, 2107-2138.

Diamond, D.W., (1989), ”Reputation Acquisition in Debt markets”, *Journal of Political Economy*, 98, 828-862

Diamond, D.W., and R. Rajan, (1998), ”Liquidity Risk, Liquidity Creation and Financial Fragility: A Theory of Banking”, mimeo, University of Chicago.

Diamond, D.W., and R. Rajan, (1999), ”A Theory of Bank Capital”, mimeo, University of Chicago.

Goldsmith, R., (1969), Financial Structure and Development, Yale University Press, New Haven.

Goldsmith, R., (1985), Comparative National Balance Sheets, University of Chicago Press, Chicago.

Gompers, P., (1995), ”Optimal Investment, monitoring, and the staging of venture capital”, *Journal of Financial Economics*, 50, 1461-1489.

Houston, J. and C. James, (1996), ”Bank Information Monopolies and the Mix of Private and Public debt Claims”, *Journal of Finance*, vol. 51, 1863-1890.

Hoshi, T., A. Kashyap, and D. Scharfstein, (1990a), ”The Role of Banks in Reducing the Costs of Financial Distress in Japan”, *Journal of Financial Economics*, 27, 67-88.

Hoshi, T., A. Kashyap, and D. Scharfstein, (1990b), ”Bank monitoring and investment: evidence from the changing structure of Japanese corporate banking relationships”, in R. Glenn Hubbard ed.: *Asymmetric Information, Corporate Finance and Investment*, (University of Chicago Press, Chicago, IL).

Hoshi, T., A. Kashyap, and D. Scharfstein, (1991), ”Corporate Structure, Liquidity, and Investment: Evidence from Japanese Panel Data”, *Quarterly Journal of Economics*, 27, 33-60.

Jayaratne, J. and P. E. Strahan, (1996), ”The Finance-Growth Nexus: Evidence from Bank Branch Deregulation”, *Quarterly Journal of Economics*, CXI, 639-671.

Jensen, M., and W. Meckling, (1976), ”Theory of the Firm: Managerial Behavior, Agency Costs and Capital Structure.” *Journal of Financial Economics*, 3, 305-360.

King, R. and R. Levine, (1993), ”Finance and Growth: Schumpeter Might Be Right”, *The Quarterly Journal of Economics*, vol. CVIII, No. 3, 681-737.

Kreps, D. and R. Wilson, (1982), ”Reputation and Imperfect Competition”, *Journal of Economic Theory*, 27, 253-279.

Kukies, J. (in progress), CRSP paper, University of Chicago.

Kumar, K., R. Rajan and L. Zingales, (1999), ”What are the determinants of firm size?”, mimeo, University of Chicago.

La Porta, R., F. Lopez de Silanes, A. Shleifer, and R. Vishny, (1997), ”The Legal Determinants of External Finance”, *Journal of Finance*, 52.

La Porta, R., F. Lopez de Silanes, A. Shleifer, and R. Vishny, (1998), ”Law and Finance”, *Journal of Political Economy*, 106, 1113-1155.

Levine, R., (1997), ”Financial development and Economic Growth: Views and Agenda”, *Journal of Economic Literature*, vol. 35, 688-726.

Levine, R., (1999), ”Law, Finance and Economic Growth”, forthcoming, *Journal of Financial Intermediation*.

Levine, R. and S. Zervos, (1998), ”Stock Markets, Banks, and Economic Growth”, *American Economic review*, vol. 88, No. 3, 537-558.

Lucas, R. E., (1988), ”On the Mechanics of Economic Development”, *Journal of Monetary Economics*, 22(1), 3-42.

Merton, R.C. and Z. Bodie, (1995), ”A conceptual framework for analyzing the financial environement”, in The Global Financial System: A functional Perspective, Dwight Crane ed., Boston MA, Harvard Business School Press.

Myers, S., (1977), ”The determinants of corporate borrowing”, *Journal of Financial Economics*, 5, 147-175.

Peach, W. N., (1941), *The Security Affiliates of National Banks*, Baltimore: Johns Hopkins Press.

Peek, J., and E. Rosengren, (1998), ”The International Transmission of Financial Shocks: The Case of Japan”, *American Economic Review*, 87, 495-505.

Petersen, M. and R. Rajan, (1995), ”The Effect of Credit Market Competition on Lending Relationships”, *Quarterly Journal of Economics*, vol. 110, 407-443.

Rajan, R., (1998), ”The Past and Future of Commercial Banking Viewed through an Incomplete Contract Lens”, *Journal of Money, Credit, and Banking*, vol. 30: (3), 524-550.

Rajan, R. and H. Servaes, (1997), ”Analyst Following of Initial Public Offerings”, *Journal of Finance*, vol. 52, 2, 507-530.

Rajan, R. and L. Zingales, (1995), ”Is There an Optimal Capital Structure? Some Evidence from International Data”, *Journal of Finance*, vol. 50, 1421-1460.

Rajan R. and L. Zingales, (1998a), ”The Firm as a Dedicated Hierarchy”, mimeo, University of Chicago.

Rajan R. and L. Zingales, (1998b), ”Financial Dependence and Growth”, *American Economic Review*, vol. 88, 559-586.

Rajan R. and L. Zingales, (1998c), ”Which Capitalism? Lessons from the East Asian Crisis”, *Journal of Applied Corporate Finance*, Fall 1998, vol. 11, No. 3, 40-48.

Rajan, Raghuram and Luigi Zingales, (2000), ”Politics, Law, and Financial Development”, Supplement No. 24 to the Government Inquiry on the International Competitiveness of the Swedish Financial Sector.

Robinson, J., (1952), ”The Generalization of the General Theory”, in The Rate of Interest and Other Essays, London, Macmillan, 67-142.

Schumpeter, J. A., (1911), A Theory of Economic Development, Harvard University Press: Cambridge, MA.

Shleifer, A. and R. Vishny, (1992), Liquidation value and debt capacity: A market equilibrium approach, Journal of Finance, 47, 1343-1366.

Weinstein, D. and Y. Yafeh, (1998), ”On the Costs of a Bank Centered Financial System: Evidence from the Changing Main Bank Relations in Japan”, *Journal of Finance*, 53, 635-672.

Politics, Law and Financial Development

by

Raghuram G. Rajan [*]

and

Luigi Zingales [**]

Supplement number 24 to the Government Inquiry on the International Competitiveness of the Swedish Financial Sector.

October, 1999

[*] Joseph L. Gidwitz Professor of Finance, University of Chicago, Financial Management Area, Graduate School of Business. 1101 East 58th Street, Chicago, IL 60637, USA. E-mail: raghuram.rajan@gsb.uchicago.edu.

[**] Professor of Finance, University of Chicago, Financial Management Area, Graduate School of Business. 1101 East 58th Street, Chicago, IL 60637, USA. E-mail: luigi.zingales@gsb.uchicago.edu.

Table of Contents

Kort sammanfattning (in Swedish)

Forskning under senare år visar att utvecklingen av ett lands finansiella marknader och institutioner tydligt påverkar den ekonomiska tillväxten i en positiv riktning. Frågan är då varför inte de finansiella marknader och institutioner som ett land behöver uppstår automatiskt. Historien visar att de institutionella hindren för ekonomisk utveckling nästan undantagslöst är politiska. Centraliserad och oinskränkt politisk makt har alltid hämmat ekonomisk utveckling. Detta är en förklaring till varför den finansiella utvecklingen inte har skett i jämn takt utan med språng, där politiska händelser som krig och revolutioner har hämmat eller t.o.m. reverserat utvecklingen. Ekonomiska marknader nådde i många länder en höjdpunkt strax före första världskriget, kollapsade med kriget och dess följder, och har återhämtat sig först under det senaste årtiondet. Det juridiska systemets utformning i dessa länder har varit en orsak till utvecklingen, dock endast en bidragande orsak.

Abstract

In recent years, research has shown that the development of a country's financial markets and institutions contribute substantially to its subsequent economic growth. But why does finance not develop to meet a country's needs? History suggests that the institutional impediment to financial development is almost always political. In particular, centralized and unconstrained political power has invariably retarded financial development. This can explain why financial development has not progressed smoothly but has had ups and downs, buffeted by political events such as war and revolution. Financial markets in many countries reached their pinnacle just before the First World War, collapsed with the war and its aftermath, and have reemerged only in the last decade. The nature of the legal system in these countries plays only a supporting role in explaining what happened.

Introduction[1]

The development of a country's financial markets and institutions dramatically increases the growth of industries, such as Computers or Pharmaceuticals, that need a lot of external, long-term finance.[2] More interestingly, financial development leads to more firms being born in industries whose assets are largely long term, or in industries where assets are primarily intangible ones such as R&D.[3] Thus financial development facilitates the spread of new ideas and leads to the "creative destruction" of the old. By freeing those with ideas from the need for personal wealth to finance them, financial development creates equality of opportunity for everyone in society, something that hitherto has only been a dream.

But how can an economy reach this utopia? In other words, what drives the development of financial markets and institutions? The simple answer, and one favored by many economists, is demand. When opportunities arise in an economy that require financing, the economy will develop the necessary markets and institutions to finance these opportunities. For example, the enormous financing requirements of railroads in the United States (one billion dollars up to 1867 and 10 billion up to 1890) lead to the development of public markets for corporate debt and later for stock, with 40 % of this capital coming from Europe.[4] Financial institutions such as investment banks, including the famous Morgan bank, emerged to underwrite and distribute these securities and to reassure European investors that the money was properly invested. Thus the needs of the railroads lead to the creation of financial infrastructure in the United States that was then available to finance other industries that came later. What we have just described is nothing but the reverse of Say's Law – demand creates its own supply.

Unfortunately, this is, almost always, an oversimplification. It cannot explain why countries at a similar level of *economic* development differ

[1] Prepared for the Symposium on the International Competitiveness of the Swedish Financial Industry organized by the Bank of Sweden Tercentenary Foundation. We benefitted from conversations with Randy Kroszner, Kristian Rydqvist, Andrei Schleifer, and Rob Vishny.

[2] See Rajan and Zingales (1998a).

[3] See Rajan and Zingales (2000).

[4] See Engelbourg and Bushkoff (1996) and Chandler (1990).

so much in the level of their *financial* development. While there is great controversy on how to measure financial development properly, there is little disagreement that, overall, Anglo-American countries score better on most measures than other, similarly developed countries, like France, Germany, and Italy.[5] Perhaps these countries do not need a system like the Anglo-American ones. Yet the desire for countries all over the world to imitate Anglo-American institutions such as venture capital, transparent accounting standards, and liquid equity markets suggests either that investment opportunities all over the world have become similar to what Anglo-American countries faced in the past, or that cross-border competition in financial and product markets is forcing financial systems to evolve towards more efficient structures. The latter explanation seems more plausible; The implication is there are impediments to financial development that prevent the demand for finance from being met efficiently.

The failure of our version of Say's law could be attributed to four, non-mutually exclusive, factors. First, there might be significant fixed costs that need to be covered before any financial intermediary can operate. Not surprisingly, in the Middle Ages money changers (the precursors of modern banks) were concentrated only in the major cities or where seasonal fairs took place, because only in these locations was business volume sufficient to pay for the fixed cost of operating. For similar reasons, some small rural towns even today have no banks.

We will argue that this factor should not be ignored in underdeveloped economies or geographically isolated communities, but it is not a satisfactory explanation for the differences in financial development across countries. It is hard to argue that financial systems in Continental Europe today are underdeveloped relative to Anglo-American ones because the amount of funds that need to be intermediated in Continental Europe is not sufficient to pay for the initial set-up costs.

A second impediment in the way of finance responding to economic needs is the time required to acquire knowledge and develop certain skills. For example, most banks today extend credit to their clientele on the basis of formal models of the credit worthiness of the borrowers. These models are based on statistical estimates of the influence various indicators have in predicting default. Thus, they can only be adopted when there exists a large enough data set with the borrower characteristics and default histories, so that financial intermediaries can compute reliable estimates. Obviously, these data and the accompanying evaluation expertise cannot be created overnight, but require time to

[5] See La Porta et al. (1997).

build. Another factor that requires time to build is reputation. Due to the difficulties in writing and enforcing fully comprehensive contracts, the reputation of the parties involved becomes essential in smoothing the working of financial markets. Reputations cannot be acquired overnight: they can only be built slowly.

The time-to-build argument can certainly explain a lag between the development of the need for finance and the emergence of a system able to fully satisfy those needs. But this friction alone cannot explain why Continental European financial systems are less developed than Anglo-American ones. In both sets of countries the need for finance arose long time ago, and any difference in the set of skills and knowledge initially available should have long disappeared.

Even after the set-up costs are incurred, a financial market needs liquidity – the ability for an agent (a seller or borrower) to get full value for her goods. A number of factors determine a market's liquidity but perhaps most important is that a large number of persons think of that particular market as the natural location for their trades. Thus there is a chicken and egg problem with liquidity – people will not trade in a particular market unless they think the market is liquid, but the market will not be liquid unless they trade. Liquidity reasons alone cannot explain the emergence of a financial center, but they can explain why an established financial center can survive after the reasons that spurt its birth have vanished. An established financial center attracts traders, because they anticipate many other traders will go there (it becomes a focal point). This can easily explain why London is still so prominent, but it can hardly explain why, for instance, in Chile the importance of the equity market is much larger than in Austria (a country with a per capita income almost six times as big).

The third potential explanation for why the demand for financing does not automatically generate its own supply is the need for an enabling infrastructure. In a financial contract, the parties exchange money for the promise of more money in the future. The value of this *promise*, however, depends crucially on the legal environment. The legal environment affects the possibility of writing certain contracts, the precision and speed with which these contracts are enforced by courts, and the punishment imposed in case these contracts are breached. For example, in the 18th century it was impossible for private parties to create a limited liability corporation without the consent of the Crown. Even today certain legal systems do not recognize or enforce certain contracts. For example, the possibility of obtaining a loan secured by the value of the entire company is a standard contract in England (called floating charge), but it is not enforced in the United States. Similarly,

most Continental European codes ban legal fees contingent on outcome, which are common practice in the United States.

These differences in legal system might matter. Recent influential research by La Porta et al (1997, 1998) finds countries belonging to the Common Law tradition have more statutes protecting investors, and more developed financial markets. But if certain norms or statutes are better in promoting financial development, why do other countries not simply adopt them? For instance, while joint stock companies were first allowed to incorporate freely in England in 1853, the concept was quickly imitated in France, where laws passed in 1863 and 1867 allowed free incorporation. Thus, while different legal systems might have a different propensity to generate legal and contractual innovation, in principle there is no restriction to imitation, eliminating the possibility of enduring differences in the level of financial development. In fact, we will show that with the exception of England, the most developed countries in the world had a similar level of financial market development in 1913. The differences in the legal system existed then also, suggesting that if they did not create differences in financial development then, we should not attribute differences in financial development between countries today solely to differences in legal system.

This leads us to the last possible cause of financial backwardness: the lack of political support in favor of financial development. Governments have great influence on financial markets. Excessive regulation and taxation (or worse expropriation) can kill a market, as can the lack of just and swift enforcement of contracts. The government also has an important role in designing and enforcing laws. Thus, it is obvious that the lack of political support can seriously jeopardize financial development. But why do some political systems foster financial development, while others do not?

Before answering this question, we need to review the working of the political system and its two main players: the main interest groups and the Government. Interest groups pressure the government to favor them. The nature of the government – whether democratic, oligarchic or a monarchy – will determine which group has more saying and to what extent its wishes get translated into actions. But the government does not simply reflect the wishes of the dominant interest group, it also has its own agenda. This is clearly the case in a monarchy, but also in a democracy where the entrenched government bureaucracy has its own perks and survival at stake. Thus, asking why some political systems oppose financial development is tantamount to asking why some influential groups or the Government (or both) oppose financial development.

There are certainly groups that win and those that lose from financial development. In a primitive financial system, the only people who can take advantage of profitable new opportunities are those that have the financial resources to do so, or can easily gather them using as collateral their stock of initial wealth. These sources of Old Wealth have an incentive to prevent competition by opposing financial development. Moreover, even if the Old Wealth is incapable of taking advantage of the opportunities, it has an incentive to prevent others from doing so since its relative position and power is much better in an underdeveloped economy, and the resources it can command greater. Thus, for example, the land-owning gentry invariably opposed financial development.

By contrast, the industrial bourgeoisie has strong incentives to press for financial development, because of the tremendous opportunities it affords them to improve their lot. This may explain why the countries of Continental Europe developed vibrant financial markets only after the bourgeois revolutions of 1848 and beyond. Russia, on the other hand, never had such a revolution and, probably as a result, had the most underdeveloped markets in Europe. As Gerschenkron (1962) argues, the Russian elites actively opposed financial development and it was the state that had to finance large investment such as the railways. Thus, the exent to which the dominant elite support or oppose financial development crucially depends on the nature of the dominant elite and of the opportunities it faces.

Also the Government itself can be in favor or against financial development, depending on the circumstances. The opposition arises for two reasons. First, the emergence of vibrant financial markets and institutions requires constitutional limitations on the arbitrary powers of government, in particular to ensure that investors' property rights are protected against expropriation. These limitations may impede government ambitions in other areas. Moreover, even when governments do become law-bound, they have a better ability to direct finance when the financial system consists of cartelized institutions than when it consists largely of securities markets. Thus they have an incentive to favor institutionalized finance over markets. For example, the militaristic Japanese government in the 1930s suppressed smaller banks and forced them to merge with large banks in an effort to direct finance towards armaments, thus initiating the Main bank system.[6] The hitherto vibrant arm's length markets were also suppressed; banks, with the acquiescence of the Ministry of Finance, formed a Bond Committee in 1933 which determined which firms could issue bonds, on what terms, and when. Clearly, banks had little incentive to allow firms to finance in the bond

[6] See Ueda (1994).

market. Of course, government can also be a force for the development of financial system if it is also in its own interest.

Finally, the ability of the dominant elite and the Government to affect the final outcome is not independent of the legal and political system. A political system where power is more concentrated makes it easier for a Government to act swiftly. Similarly, a legal system where the legislator makes the law, as in the civil law system, facilitates Government intervention, while a legal system based on precedents, like the common law one, makes it more difficult for the Government to alter the status quo. Thus, the political will of the Government and/or the dominant elite does not translate into action in the same way in different legal environments. Actions, whether favorable to or against financial development, will bear fruit more quickly in more centralized Governments and in civil law countries.

Can this theory explain the historical evolution of financial markets? It certainly is consistent with the broad outlines. Consistent with the political theory, all the early financial centers of the Modern world, starting with Florence and Venice, moving on to Hamburg and the cities of the Hanseatic League, and culminating in Amsterdam, emerged in little independent political entities controlled by the bourgeoisie. None of the political centers of big and powerful centralized states (Paris, Madrid, or Vienna) emerged early on as important financial centers. Also the emergence of London can be traced to the fortuitous coincidence of the two crucial catalysts predicted by the theory: an elite who could benefit from financial development and a weak Government. In England, as in all the other States, the dominant elite was the land aristocracy (gentry). But unlike other States the English land aristocracy in the 17th century could benefit from a development of the capital markets. About the same time, the English crown, which had lost its credibility by expropriating bankers in the famous Stop on the Exchequer (1672), was forced to introduce reforms that facilitated its own borrowing, and coincidentally, financial development. The result was that by the end of the 18th century London emerged as the financial capital of the world.

This predominance of London carried well into the 20th century. But during the second half of the 19th century, other financial centers started to emerge: not only New York, but also Paris and, to a less extent, Berlin. In fact, in 1907 the total amount of securities traded in Paris was slightly larger than that traded in New York, despite the fact that French GNP was only one fourth of U.S. GNP.[7] Interestingly, 58 % of the securities traded in Paris were foreign, while almost no foreign securities

[7] Davis and Neal (1998).

traded in the U.S. Thus, in spite of a slow start, Paris had become the second financial capital of the world. What explains this catch up?

The second half of the nineteenth century saw, according to Eric Hobbsbawm, the Age of Capitalism and the triumph of pro market forces. The revolutions following 1848 put the bourgeoisie in power in most European states. The prolonged period of peace in Europe, between 1870 and 1914, brought economic interests to the fore. The major countries of Europe went on the Gold Standard to facilitate cross-border trade and international capital flows. Governments, even if not propelled by their bourgeois supporters, found it in their interest to aid the process of integration because it enhanced their revenue and financing options. In fact, the breathless prose of learned commentators in 1913 is reminiscent of what one hears today:

> "The world is becoming one in an altogether new sense…As the earth has been narrowed through the new forces science has placed at our disposal…the movements of politics, of economics, and of thought, in each of its regions, become more closely interwoven…Whatever happens in any part of the globe has now a significance for every other part. World History is tending to become One History."[8]

With the industrial bourgeoisie in power, Governments throughout Europe promoted financial development. Where legislative power was more concentrated at the top, as in civil law countries, this promarket intervention could be more effective. While it took over a century and a half for the English common law system to work out the limited liability form to its satisfaction, ten years were sufficient for the French civil code to imitate it. This explains the almost instant success of Continental European Governments in promoting financial development.

But World War I and its aftermath changed everything. Financial markets declined significantly in many countries between 1913 and 1948 even as a proportion of GDP, and in some cases have not reverted back to their prewar level even today. In Italy, for example, the number of publicly traded companies was higher in 1913 than in 1985.[9] Interestingly, this decline was experienced not only by countries that lost one or both wars, but also by countries (like France) that won both, and even by countries, like Sweden, that did not participate in either. What caused the retreat of financial markets? Why was it more pronounced in some countries than in others?

[8] Address on April 3, 1913 of Mr Bryce, President of the International Congress of Historical Studies, cited in Powell (1915), p. 704.
[9] Pagano, Panetta, and Zingales (1998).

In our view, there were two main reasons; one a bottomup upwelling of the desire for insurance, the other, a top down desire for national security that required the coordination of national resources. These political forces managed to assert themselves, especially in economies with centralizing tendencies. Since insurance and coordination of the kind desired are not easily supplied by the market, the market was forced to retreat in country after country, and supplanted by hierarchical organizational structures.

Let us explain in greater detail. The expansion of markets creates risks, which are not easily insured. For example, the lowering of trade barriers by an economy motivated by international trade could lead to entire industries becoming uncompetitive and the human capital of their workers redundant. It is hard to insure against this risk because the risk is hard to diversify away, and also insurance breeds standard problems of moral hazard and adverse selection.

Similarly, the existence of competitive arm's length securities markets makes it harder for long term bank-firm relationships to form, and increases the risks to firms of being shut off from credit. The rationale is that these relationships are a form of insurance; a firm in trouble gets credit at a below-competitive rate in return for which it pays the premium by giving more of its business to the relationship bank in good times. A more competitive outside environment puts great stress on a relationship because opportunities outside the relationship become more attractive. It becomes much harder for either party to feel confidant in the give-and-take that is necessary for the relationship to work becomes the other party could just leave after taking. Since a relationship is governed by mutual confidence rather than contracts, an increase in competition in financial markets leads to a breakdown in confidence, and the quick demise of relationships founded on nothing more substantial than the absence of alternatives.[10]

More generally, as first suggested by Polanyi (1944), competitive markets whether in labor, industry, or finance tend to put strains on social relationships, which are based on implicit contracts. These social relationships are the primary source of insurance before the advent of the market. Moreover, markets tend to accentuate the difference between the incompetent, the unskilled, or the untalented and the more qualified, thus exacerbating the need for insurance. Thus competitive markets destroy old sources of insurance while creating new needs for it. Unfortunately, competition again – when coupled with the lack of commitment that

[10] This idea was initially proposed by C.Mayer "New Issues in Corporate Finance", *European Economic Review*, 32 (1988), p. 1167-1189. Petersen and Rajan (1995) model it formally and find empirical evidence for it.

leads to incomplete contracting or free-riding – makes it hard for markets to provide the necessary cross-subsidies that mitigate its harshness.[11]

In good times, the greater allocative efficiency of the market papers over the lack of insurance. Everyone benefits from the rise in incomes. But in major downturns, the political pressure from the large masses of losers becomes hard to resist. In particular, the post-war unemployment, financial crises, and eventually the Great Depression – that could all be traced back to the disequilibrating effects of the World War – stood in direct contrast to the full employment, productive efficiency, and stability that most European economies enjoyed under the command structure necessary for military production during the war. To the common man there seemed something wrong with the market economy and these beliefs fuelled the antimarket revolt that Polanyi called the Great Transformation.

In order to provide insurance in the form of higher employment, higher prices or wages to certain sectors, and financial sector stability, governments had to intervene massively in the working of markets – be they financial, labor or trade. But these typically distortionary actions were hard to undertake while the economy remained open to international competition. Protectionist tendencies therefore came to the fore. As more and more countries closed themselves off from the international economy, and went off the Gold Standard so as to better effect internal cross-subsidies, the benefit to any country of resisting internal pressure and maintaining open markets diminished. International markets closed down.

The anti-market reaction took different forms. In Italy, the Fascists started the first major wave of nationalization of industry. In Argentina, the populist movement of Juan Peron handed out large subsidies. In Sweden, the 60-year domination of the Labor party, which created the Swedish welfare state, began. In England, the rate set by the ostensibly independent Bank of England stood at a political 2 percent between June 1932 and the outbreak of the war. Even the United States, traditional paladin of the free market economy, had its New Deal under Roosevelt.

The intensity of this anti-market revolt varied country by country, depending on the severity of the economic shock, its exposure to those rearming, and most importantly, the country's inherent centralizing tendencies. In particular, we will argue that the different legal and institutional environments provided widely varying degrees of protection against the anti-market backlash. Countries with a civil law system,

[11] For models showing the difficulties of cross-subsidization in a market environment, see Diamond and Dybvig (1983) and Jacklin (1986).

where the legislature makes the law, became an easy prey for the political movements advocating a command and control system. Common Law provided a stronger bulwark against centralization. First, the English common law tradition emphasizes restraining the government and protecting the individual, making more difficult the centralization of the economy. Second, Common Law is based primarily on judicial precedent, rather than upon the act of a legislator. Thus, a new political climate is less able to transform the structure of society. Third, even though Common Law is hard to change from the center, it evolves at the periphery, and innovates around legislative or administrative roadblocks set up by the center. In our view the decentralized nature of the Common Law system, and its ability to resist anti-market forces can explain the continued vibrancy of financial markets in Common Law countries even when markets elsewhere declined.

Our analysis emphasizes the importance that political consensus (or lack of) plays in the development and defense of markets in general and financial markets in particular. Its implications are not that all countries that did not have the good ”fortune” of being colonized by the British should adopt British common law. Rather, it is a note of caution against taking for granted that financial development is unidirectional, and that the generalized consensus in favor of markets is irreversible because of their obvious efficiency. History shows that in time of crisis a political backlash may occur and it can have very long-term consequences. As a result, any policy designed to promote financial development cannot ignore its political effects. Only if the reforms can enhance popular consensus for markets can they have positive long-lasting effects.

To summarize our main thesis:

1) Financial development results when the politically powerful develop a consensus in its favor.
2) The centralization of political power in an economy tends to magnify the interests of those in power. This can speed up financial development when the center is in favor of markets – an all too rare occurrence – and reverse it when it is not.
3) Decentralization of political power, coupled with an efficient legal system tends to be most conducive for the protection of property rights and the financial innovation necessary for financial development. Decentralization also ensures that financial development is hard to reverse when the consensus at the center changes.
4) A Common Law system allows for more contractual and legal innovation, which makes it easier to by pass the potential roadblock imposed by an elite that opposes financial development. Hence it is

> more conducive to financial development. But its innovations can easily be imitated by a suitably motivated Civil law system. Thus, the greater financial development in Common Law countries is not because laws are better in those countries, but because the decentralization of Common Law makes it easier for financial markets to develop in spite of political opposition and makes it harder to reverse this development in the face of political challenges.

Now that we have laid down the main arguments, we move to explore them in greater detail. In the first section we present some aggregate data on the differences in financial development across countries. Section 2 briefly reviews the fixed costs and learning-by-doing theories of financial backwardness. Section 3 reviews the nature of different legal systems. Section 4 explores the possible theoretical links between legal system and financial development. Section 5 tests the predictions of the legal theories of financial development by comparing the level of financial development circa 1913. Section 6 proposes a political theory of why and when the dominant elite and the Government will oppose the development of an efficient financial system. Section 7 analyzes how historically financial systems emerged in spite of this resistance. Section 8 tries to identify why the evolution of finance in the last ninety years has differed so greatly across countries. Section 9 discusses the policy implication of our analysis. Section 10 concludes.

1 Financial development across countries

There is no consensus on the correct method to measure financial development. Theoretically, the right measure would be the ease with which companies in need of external funds can access them and the premium they have to pay for these funds. Unfortunately, these are not measures that can be easily computed even in the most developed countries, let alone for a large cross section of countries.

Thus, the standard practice, which we will follow here, is to compute ratios of the size of different components of the financial system to measures of the size of the economy. Table 1 reports such indicators for all the countries that have a stock exchange listed with the *Federation International Bourse Valeurs (FIBV),* the international association of Stock Exchanges.[12] Even though this is a fairly comprehensive list, which includes a lot of developing and former socialist economies, it is important to note that countries enter the sample only if they have a stock exchange.

For ease of comparison, the countries are ordered in terms of per capita GDP at the end of 1996. The first measure reported is stock market capitalization over GDP. While there is a positive correlation between relative size of the stock market and the level of economic development (measured as per capita income), this association is far from overwhelming. Germany and the Netherlands are two countries that are very similar in every dimension (including the level of per capita GDP), but in the Netherlands the equity market capitalization is almost as much as the GDP, while in Germany only 30 %. The same can be said for the Philippines and Indonesia. Thus, the size of the equity market does not seem to be a simple reflection of the level of economic development.

[12] We omit information for Azerbaijan, Croatia, Ivory Coast, Jugoslavia, Lebanon, Luxembourg, Hong Kong, and Taiwan because the International Financial Statistics do not report GNP data for these countries.

Table 1. Financial and Economic Development around the World 1996, end of year data

Country	Per capita GDP	Equity market over GDP	Bond Market GDP	Domestic Credit over GDP	Total over GDP	Equity raised over GCF	Bond raised over GCF	Total over GCF
Switzerland	38,147	1.48	0.37	1.67	3.52	NA	0.25	NA
Norway	36,047	0.36	0.14	0.59	1.09	0.04	0.08	0.12
Japan	34,434	0.72	0.04	1.15	1.91	NA	NA	NA
Denmark	34,062	0.40	0.99	0.31	1.70	0.03	0.63	0.66
USA	28,766	1.11	0.03	0.45	1.59	0.17	0.60	0.76
Sweden	27,833	0.98	0.45	0.37	1.80	0.09	0.32	0.42
Austria	26,702	0.16	0.19	1.01	1.36	0.01	0.09	0.10
France	25,712	0.39	0.11	0.82	1.32	0.05	0.10	0.15
Belgium	25,664	0.46	0.01	0.66	1.14	0.06	0.02	0.09
Singapore	25,317	1.62	NA	0.96	NA	NA	NA	NA
Netherlands	24,464	0.99	0.35	1.06	2.40	0.18	0.29	0.47
Germany	23,151	0.35	0.59	1.26	2.20	0.04	NA	NA
Australia	21,705	0.78	0.00	0.78	1.56	0.15	NA	NA
Finland	21,651	0.56	0.08	0.68	1.32	0.03	0.09	0.13
UK	21,348	1.31	0.22	1.23	2.76	0.17	0.33	0.50
Italy	21,335	0.21	0.03	0.51	0.75	0.02	0.05	0.07
Ireland	20,051	0.49	NA	0.36	NA	NA	0.11	NA
Canada	19,978	0.81	0.00	0.47	1.28	0.09	0.00	0.09
New Zealand	18,963	0.54	NA	0.92	NA	0.10	0.02	0.11
Kuwait	17,670	0.61	NA	0.34	NA	1.18	NA	NA
Israel	16,347	0.37	0.03	0.72	1.13	0.03	0.00	0.03
Spain	14,398	0.43	0.02	0.71	1.16	NA	0.05	NA
Greece	11,505	0.20	0.01	0.23	0.44	0.03	0.03	0.05
Portugal	10,758	0.23	0.06	0.67	0.96	0.03	NA	NA
Korea	10,146	0.30	0.19	0.64	1.13	0.04	0.22	0.26
Slovenia	9,142	0.05	0.00	0.29	0.34	0.09	0.01	0.10
Bahrain	8,908	0.94	NA	0.97	NA	0.04	NA	NA
Argentina	8,448	0.15	0.01	0.18	0.34	NA	0.06	NA
Oman	7,182	0.20	NA	0.26	NA	0.10	NA	NA
Uruguay	5,476	0.02	NA	0.29	NA	0.00	NA	NA
Czech Republic	5,012	0.28	NA	0.62	NA	0.15	NA	NA
Brazil	4,742	0.29	0.00	0.26	0.55	0.08	NA	NA
Chile	4,672	0.98	0.18	0.56	1.72	0.14	0.15	0.29
Malaysia	4,666	3.10	0.03	0.93	4.07	0.14	0.04	0.18
Hungary	4,158	0.12	NA	0.18	NA	0.05	NA	NA
Slovakia	3,362	0.07	NA	0.32	NA	0.00	NA	NA
Mexico	3,353	0.33	0.02	0.18	0.53	0.01	0.02	0.04
Poland	3,278	0.07	0.00	0.16	0.23	0.01	0.00	0.01

Country	Per capita GDP	Equity market over GDP	Bond Market GDP	Domestic Credit over GDP	Total over GDP	Equity raised over GCF	Bond raised over GCF	Total over GCF
Panama	3,034	0.16	NA	0.78	NA	0.00	NA	NA
Thailand	2,990	0.53	0.00	1.02	1.55	0.06	0.00	0.06
Russia	2,742	0.00	NA	0.07	NA	NA	NA	NA
South Africa	2,736	2.06	0.01	0.70	2.77	0.30	0.00	0.30
Venezuela	2,714	0.16	NA	0.08	NA	NA	NA	NA
Botswana	2,686	0.08	NA	0.11	NA	NA	NA	NA
Jamaica	2,557	0.30	NA	0.24	NA	NA	NA	NA
Costa Rica	2,505	0.09	NA	0.18	NA	NA	NA	NA
Colombia	2,503	0.15	NA	0.21	NA	0.05	NA	NA
Peru	2,407	0.22	0.02	0.19	0.43	0.02	0.04	0.06
Iran	2,179	0.10	NA	0.18	NA	0.03	0.00	0.03
Lithuania	2,127	0.16	NA	0.11	NA	NA	NA	NA
Turkey	2,118	0.23	0.00	0.22	0.45	0.04	0.02	0.06
Paraguay	1,889	0.04	NA	0.23	NA	NA	NA	NA
Tunisia	1,542	0.27	NA	0.67	NA	0.07	NA	NA
Equador	1,435	0.13	NA	0.27	NA	NA	NA	NA
Marocco	1,321	0.24	NA	0.28	NA	0.09	NA	NA
Jordan	1,299	0.67	NA	0.65	NA	0.11	NA	NA
Swaziland	1,209	0.41	NA	NA	NA	NA	NA	NA
Latvia	1,209	0.05	NA	0.12	NA	NA	NA	NA
Philippines	1,162	0.96	0.00	0.48	1.45	0.08	0.00	0.08
Indonesia	1,137	0.40	0.00	0.55	0.96	0.08	0.00	0.08
Namibia	1,113	0.27	NA	0.69	NA	3.58	NA	NA
Egypt	1,108	0.21	NA	0.37	NA	NA	NA	NA
Sri Lanka	738	0.14	NA	0.25	NA	0.04	NA	NA
Zimbabwe	664	0.46	NA	0.21	NA	0.06	NA	NA
China	663	0.08	NA	0.94	NA	NA	NA	NA
Honduras	610	NA	NA	0.23	NA	NA	NA	NA
Bulgaria	428	0.00	NA	0.36	NA	0.00	NA	NA
Pakistan	404	0.22	NA	0.25	NA	0.07	NA	NA
Nigeria	383	0.31	NA	0.12	NA	0.08	NA	NA
India	380	0.34	NA	0.26	NA	0.11	NA	NA
Zambia	373	0.07	NA	0.09	NA	NA	NA	NA
Ghana	346	0.27	NA	0.06	NA	0.07	NA	NA
Kenya	296	0.19	NA	0.28	NA	0.05	NA	NA
Bangladesh	255	0.11	NA	0.28	NA	0.02	NA	NA

Sources: FIBV, Annual Reports and IMF: International Financial Statistics.

This is true also for the size of the bond market. While the available data from the FIBV suggests a stronger association between the level of economic development and the size of the bond market, this is not the entire story. For example, Italy and the United Kingdom have a similar per capita income, but very different bond markets (3 % of GDP for Italy, 22 % for the U.K.). Interestingly, among the countries for which data are available there is no association between the level of development of the bond market and that of the equity market. Germany, which has a relatively small equity market (35 % of GDP) has a relatively large bond market (59 %).

A third indicator of financial development, although not necessarily an indicator of development of financial markets, is the amount of domestic credit over GDP. This is a measure of the amount of credit extended by banks. Again there is a strong association between this indicator and the per capita income, as also considerable unexplained variation: Austria, which has a per capita income very similar to Sweden, has three times the level of credit per unit of GDP. Similarly, the U.K. has two and a half times the level of domestic credit as Italy. Interestingly, the level of domestic credit is positively correlated with the importance of the equity market, suggesting complementarity rather than substitution between the two. Some countries appear to be more financially developed on every dimension.

One important drawback of the above measures is that they record the stock of financial instruments outstanding, rather than the flow of investments intermediated by the financial system. This drawback is particularly serious in the case of the equity market capitalization, which measures the market value of the existing companies rather than the amount of funds raised in the equity market. Thus, the Dutch stock market appears very large because of the presence of the oil giant Royal Dutch Shell. But Royal Dutch Shell does not raise much funds externally, thus its presence in the Netherlands does not necessarily indicate that companies can raise funds easily in the Dutch equity market. A better measure, therefore, is the fraction of total investment that is financed by equity issues (the sum of equity issued in secondary markets by publicly traded companies and shares sold at the initial public offering by newly listed firms). Unfortunately, this datum is not available for many countries. For the countries for which it is available, however, it confirms the previous results. The level of financial development differs greatly across countries even if we compare countries with a similar level of economic development. Furthermore, Anglo-American countries, (United Kingdom and its former colonies) financed a much larger fraction of their investment from the market, than

the rest of the world (16 % vs. 6 % even after eliminating Namibia, which is a big outlier).

The data on funds raised through bond issues should be interpreted with caution because they are gross bond issues, that is, they do not subtract the amount of bond retirement. Nevertheless, they seem to confirm the overall pattern, as do the sum of the two.

In sum, the data on financial development around the world seem to reject a simple variant of Say's law. Financial development does not go hand-in-hand with economic development. In fact, all the different measures of financial development indicate wide differences across countries with a very similar level of economic development. Why is this the case? This is the question we will try to address in what follows.

2 Fixed costs and time-to-build theories of financial backwardness

Clearly, there are large fixed costs in setting up financial infrastructure.[13] Furthermore, it may be hard for private enterprise to internalize all the gains to doing so, especially if these gains are spread over a long period of time. If there is no preexisting financial infrastructure, the economy needs either the fortuitous combination of rich individuals with the will at a time when there is a sudden large demand for financing (or a farsighted government that can pool resources) for the infrastructure to be set up. Chandler (1977) suggests this is, in fact, what happened in the United States. The financial sector, especially investment banks and the corporate bond market, developed to meet the enormous financing needs of railroads in the midnineteenth century. The total investment in American railroads came to a billion dollars in 1867, three billion in 1873, five billion in 1880, and fully ten billion in 1890.[14] This was far in excess of the financing ability of a young nation, especially given the enormous financial strains caused by the civil war. Therefore, the source for over forty percent of this capital before 1914 was Europe.

At that time, capital in Europe could be obtained for about half its cost in the United States. But European investors were fearful due to the lack of reliable information and the fact that ”legal and quasi legal safeguards for the buyer were almost entirely lacking. Recourse to American courts would hardly help investors in far-off Europe, who were confused in any case by the differences between state and federal jurisdictions”.[15] The solution to this problem was the creation of institutional investors, who looked after the interest of British investors on the American soil. This infrastructure, built to finance railroads, was later available to meet the less demanding needs of industrial firms around the turn of the century.

Even after the set-up costs are incurred, a financial market needs liquidity – the ability for an agent (a seller or borrower) to get full value for her goods. A number of factors determine a market’s liquidity but

[13] For models on fixed costs of financial development and growth, see Bencivenga and Smith (1991) and Greenwood and Jovanovic (1990).
[14] Egelbourg and Bushkoff (1996).
[15] Egelbourg and Bushkoff (1996).

perhaps most important is that a large number of persons think of that particular market as the natural location for their trades. Thus there is a chicken and egg problem with liquidity – people will not trade in a particular market unless they think the market is liquid, but the market will not be liquid unless they trade. This again implies that the birth of a market requires the rare event that serves to coordinate expectations. For example, the United States Government issued Liberty Bonds in large numbers to finance the war effort during the First World War. People who would otherwise not touch a financial security bought these bonds for patriotic reasons. Peach (1941) argues that it was the successful experience these investors had with the bonds that made them willing to invest in arm's length securities issued by corporations, which then gave corporate securities markets liquidity, and led to the enormous expansion of these markets in the 1920s.

More generally, economic theory suggests that a market becomes liquid when, somewhat counter-intuitively, there are a large number of uninformed investors willing to grease the wheels of trade. This is because when a market consists only of informed, experienced traders, everyone is trying to second-guess the counterparty's information from the way they trade, and few trades take place. The reasoning is "if you are willing to sell me this stock and I know you are smart, I keep wondering what it is you know that I do not, and am less willing to buy". The difficulty of providing uninformed investors the confidence to participate in the market when, by their very nature, they will not collect a great deal of information is one of the most important impediments to market liquidity.

Investors obtain such confidence, not from a detailed scrutiny of the financial press but from word-of-mouth about salient events, and articles in the popular press. There is some evidence of this. Initial public offerings of stock that have higher first day returns in the United States attract more analyst following, perhaps because huge initial returns make these "story stocks" that are written about in the popular press attracting investor interest. These stocks also have higher trading volumes, a natural consequence of more uninformed interest.[16] But there is a downside to attracting the uninformed. A broad-based unpleasant experience can equally easily drive them from the market and word of mouth about such a salient event will take a long time to die down. The bottom line is that liquidity is based on past experience, takes a long time to nurture, and could be somewhat fragile for new markets.

Financial contracts also take time to perfect. There is a process of trial and error in which institutions initially, and eventually markets,

[16] See Rajan and Servaes (1997).

learn the nuances of a financial contract, how to price it, what covenants to build in, how to withstand legal scrutiny, and how to obtain effective enforcement.[17] The institutions such as exchanges, audit agencies, custodial services, rating agencies, etc., that help the smooth enforcement of arm's length contracts take time to set up, and more time to perfect their trade in local conditions.

Financial intermediaries also take time to set up and become effective. As Rajan (1998) argues, institutions intermediate where financial contracts are incomplete, either because contingencies cannot be fully foreseen in the case of innovative transactions, or because the contingencies are too complex to describe up front. The institution uses its reputation to bond itself, and its bargaining power to achieve negotiated outcomes that are superior to those that can be obtained through direct transacting in the market place. A good reputation is crucial for an intermediary because this is what convinces customers to trust it when contracts do not fully protect them, but reputations take time to build, see Kreps and Wilson (1982) and Diamond (1989), for example.[18] An institution may also be able to bond itself by putting its monetary capital or future rents at stake, but we have already argued that monetary capital may not be available in concentrated form in economies with primitive financial sectors. Moreover, even if available, monetary capital alone may not be particularly effective in providing assurances to customers, see Boot, Greenbaum, and Thakor (1993).

Finally, financial intermediaries take time to build for the same reasons that large industrial firms take time to build, see Rajan and Zingales (1999). It takes time for employees to learn to work with each

[17] See Rajan (1998).

[18] In every financial innovation, there are outcomes that are not anticipated up front, and have to be worked out amicably between intermediary and client. This is where the intermediary's reputation plays a part. Some intermediaries develop reputations for being un-cooperative, which hurts their business. An incident reported by Euromoney (April 1995, p. 35) is suggestive. In 1987, Texaco sought bankruptcy protection after being ordered to pay $ 10.5 billion in damages to Pennzoil. Even though Texaco was far from insolvent, the filing put it in technical default on a swap contract with Bankers Trust. The details of the swap contract allowed either party to walk away even if they owed money when one party defaulted (this limited-two-way-payment clause is deservedly obsolete). Bankers Trust owed $ 10 million on the swap. The expectation was that it would waive the default given that it was merely technical. Euromoney reports that "top Bankers Trust management looked at the situation, weighed the bank's relationship with Texaco – not itself known for showing much mercy to its bankers – and took the windfall gain. Texaco has refused to deal with Bankers Trust ever since".

other, to build relationships with customers, and to build routines that ensure the smooth functioning of the organization. Moreover, all this learning cannot take place instantaneously, it has to be sequenced, else it could risk overwhelming the command and control capabilities of the organization.

While these factors are clearly important in explaining what triggered the birth and early steps of a financial market, it is hardly a satisfactory explanation for today's differences in financial development across countries that we reviewed in the previous section. It is hard to argue that the French financial system today is underdeveloped relative to the English one because the amount of funds that need to be intermediated in the French economy is not sufficient to pay for the initial set-up costs. And while liquidity reasons can explain the enduring prominence of London, they can hardly explain why some late-movers progressed so rapidly and others did not. For example, in Chile the equity market capitalization is 98 % of GDP and 14 % of the funds invested (Gross Capital Formation) are raised through equity issues. By contrast, in Austria (a country with a per capita income almost six times as big), the equity market capitalization is equal to 16 % of GDP and only 1 % of the funds invested (Gross Capital Formation) are raised through equity issues.

3 The Nature of the Law

The third potential explanation for why the demand for financing does not automatically generate its own supply is the need for an enabling infrastructure. Since contracts lie at the heart of almost any financial arrangement, there is a growing consensus amongst economists that the law – which dictates permissible contracts, their interpretation, and the ease with which they can be enforced – must matter for financial development.

Clearly, specific laws make a difference to financial systems. For example, the great variety of financial institutions in the United States – ranging from traditional commercial banks to modern venture capitalists or boutiques specializing in trading derivative securities – is probably an outcome of the Glass Steagall Act of 1933, and other legislation introduced during the Great Depression. These laws, reflecting the age-old suspicion among the people of the Western frontier states of the United States of East Coast bankers, effectively curtailed the power of large financial institutions and prevented them from dominating any part of the financial sector.[19] The different legal protection granted to minority shareholders has been shown to affect the value of corporate of control.[20] Similarly, a study of G-7 countries found that perhaps the best determinant of the amount large firms borrowed was the degree to which bankruptcy laws favored creditors. Countries like the United Kingdom and Germany that had relatively severe bankruptcy laws favoring creditors were where large firms borrowed the least.[21] While these early studies suggest that laws can affect finance, they leave unanswered broader questions such as why financial markets seem to have so much easier a time emerging in some countries but not in others.

Perhaps a better starting point is not specific laws but the legal system – the way laws are formed and then enforced. Let us first describe how legal systems broadly differ across countries.

[19] See Roe (1994).
[20] Zingales (1995).
[21] See Rajan and Zingales (1995).

3.1 Legal Systems: Common Law and Civil Law

Most modern legal systems in market economies can be classified as either belonging to the Common law tradition or the Civil law tradition. The main difference between the two systems lies in who makes the law. In the Common law system, past judgements by higher courts have enormous weight. Statutes passed by legislators are, therefore, considerably refined, and even modified, by precedent set in cases by judges. By contrast, in the Civil law tradition, law is really made by legal theorists with the aim of achieving (at least theoretically) the end of justice. The proposed statutes are enacted by legislators, and become paramount. Judicial decisions do not create law, because that task, in the Civil law system, belongs solely to the legislative body. The judge tries to interpret logically the intent of the legislator in arriving at a just solution. But in practice, the courts are not completely at ease unless they can invoke one or more texts of enacted law to justify or support their judgement.[22]

While this is the main difference, there are related systematic differences in the two systems. Common Law evolved in England, and was never based on a code of laws even though the Romans had ruled England, and Roman law is the source code of most Civil codes. Over time, judges obtained enormous power because of their knowledge of precedent, and their ability to make law, and they resisted codification, which would force them to cede power to legislators and theorists. Of course, it also helped in England that while the Common law system was evolving, there were countervailing powers to the King such as rich nobles, who supported a judicial system that was not under the king's thumb. As a result, the judicial system became a bulwark against the arbitrariness of the king. This is reflected in the kinds of cases that were initially tried under the common law system – matters relating to royal finances, ownership and possession of land, and serious criminal matters affecting the peace of the kingdom.[23]

The emphasis in the Common law system is in following "due process". Justice emerges from following the formalistic principles laid down such as the selection of the correct form of action or writ by which the court can be seized, convincing the court it has jurisdiction, and following the adversarial process of argument before a jury. The English attitude is that if a fair procedure is followed, a just solution

[22] David and Brierley (1979, p. 95).
[23] David and Brierley (1979, p. 291).

automatically emerges. The emphasis on procedure is natural in a country where there was no prior body of law to refer to, and judges empirically created law.[24]

As a result, Common law is very open to any kind of question that can arise. Since it emphasizes method, answers to new questions can be found by starting with already enumerated rules and discovering the legal rule – perhaps a new one – which deserves to be applied in the specific case at hand. The facts of the case are very important, and if the judge arrives at a different decision from that suggested by precedent, it is because the facts are different.[25] Thus Common law, according to legal theorists David and Brierley, is an ”open” system.

By contrast, the Civil law system originates from the code of Justinian, laid down at a time when there was a need to obtain some uniformity in the system of justice across a very diverse empire. It is quite suited to a centralized system because any legislation modifying the law cannot be altered by judges, who instead have to find just solutions within the confines of the statutes laid down by legislators. While this makes the law more responsive to central command, it adapts much more slowly to local conditions, especially to the idiosyncrasies of the particular case. Unlike the judge under Common law who looks for factual exceptions so as to break from precedent, the judge under Civil law looks for statutes the judgement can conform to. As a result, the Civil law system is much more of a closed system.

Within the broad Romano-Germanic Civil law tradition, we can distinguish three further subcategories. The German tradition is a little more flexible than the French tradition because the German tradition allows judges to override specific legislation if it contravenes general principles of the law.[26] The French tradition, however, frowns on this, perhaps because the past French experience with unruly ”parlements” or courts of law in the 17th century convinced them that allowing judges to interpret articles would, in the end, subvert the legislative order created by the French Civil Code, largely enacted under Napoleon in the early nineteenth century.[27] In fact, the French may have been a little more convinced than others that their code was the ”perfection of reason” and the best way to reach justice was simply to follow the code.[28] Finally, the Scandinavian countries have their own Civil code which is distinct from the German or French tradition.

[24] David and Brierley (1979, p. 331).
[25] David and Brierley (1979, p. 335).
[26] David and Brierley (1979, p. 111-112).
[27] David and Brierley (1979, p. 113).
[28] David and Brierley (1979, p. 96).

In drawing out the salient features of each legal tradition, we may have caricatured them a bit. This is the price we pay for brevity. But the caricature seems to match well with differences in dominant philosophy – Common law corresponds to the empirical tradition of Locke and Hume while Civil law relates to the rationalist tradition of Descartes. Furthermore, abstraction from detail also has permitted economists to classify different countries in the world according to which tradition they most resemble. Countries adopt a particular legal system either because they developed it themselves, imitated the salient features of another country's code, or because they were conquered or colonized, and the victor passed its system on. It is useful to see how differences in legal system affect financial development, and perhaps more important, analyze why they might have an influence.

4 Does the Legal System Really Matter for Financial Development?

There are three direct channels through which a legal system can affect financial development. In what follows we will discuss the sources of these effects as well as discuss the evidence on their empirical relevance.

4.1 Law Enforcement

A financial contract is little more than a promise to repay some money in the future. The value of this promise depends crucially on the speed and efficiency with which courts provide remedies for breach. It is not only important whether a certain sum is repaid, it is also important when this sum is repaid and how reliable the expectation of repayment is. While efficient enforcement benefits all financial contracts, it benefits most arm's length contracts, which lack alternative ways of enforcement. By contrast, contracts written in the context of repeated relationships do not rely solely on the courts, and can be enforced by the threat of interrupting the relationship. Thus, inefficient enforcement should especially jeopardize the development of financial markets vis-à-vis other financial institutions like banks.[29]

There is little doubt that law enforcement is important for financial development. La Porta et al (1997) find that a country's "law and order tradition" (as compiled by *International Country Risk Guide*) contributes to explaining the relative size of its financial markets. But this is far from being the whole story. Germany, for instance, scores near the top in all the index of efficiency of the judicial system and contract enforcement. In spite of that (and of its high level of economic development), it has a fairly minuscule market. Smaller than the one of a country like Thailand, which scores near the bottom of the list in all these measures.

More importantly, contract enforcement is, by a large extent, a policy variable. Thus, an explanation of financial development based entirely on the efficiency of contract enforcement would beg the question of why some countries enforce the law better than others. An explanation in this

[29] See Rajan and Zingales (1998) for a development of this argument.

sense can only be found moving to a political view of financial development, avenue that we will pursue in Section 4.

4.2 Investor Protection

The traditional legal approach to Corporate Law was positing that contract enforcement is all that is needed for financial development. Rational investors will write in financial contracts all the necessary clauses to protect themselves. The State only role, then, would be to enforce these contracts in a prompt and just manner.

When it comes to public markets, however, two problems emerge. First, dispersed buyers find it impossible to bargain on the specific legal details of a security. They are just left with the option to buy a financial instrument ”as is” or not to buy it. In this situation, a set of prespecified rights granted by the law can reduce contracting costs. Second, dispersed investors find it costly to enforce their rights. By standardizing the rights, explicit laws may facilitate enforcement. Thus, it is possible that better laws designed to protect investors might foster financial development. The direction of causality, however, is not clear. More advanced financial markets have more experience on what it is needed to protect investors. They are, thus, more likely to pass laws in this direction.

To try to disentangle the direction of causality, La Porta et al. investigate whether the degree of investor protection differs as a function of the family of origin of a country’s legal system. Countries adopted a particular legal system centuries ago, mostly as a result of an invasion or colonization. If, more than a century later, the type of system adopted has an effect on the level of investor protection and financial development, then – the authors claim – there is a ”strong case that legal families, as expressed in the legal rules, actually affect outcomes”.[30]

In a series of influential papers, La Porta et al. show that indeed Common Law countries protect investors better. This is true not only for equity investors, but also for creditors. They then go on to relate cross-country differences in financial development with the family of origin of a country’s legal system. In discussing Table 1, we have already pointed out that the U.K and its former colonies, which are Common Law countries, have a much larger ratio of the size of stock market to GDP. The evidence presented by La Porta et al. (1997) is much more detailed. Common law countries have many more listed firms per million people and more initial public offerings per million people. For example, France

[30] La Porta, et al. (1998, p. 1126).

has 8 listed firms per million people in 1994 while Germany has only 5. By contrast, the United Kingdom has 36, the United States 30, and Israel an astonishing 128. This diversity in listed firms is also reflected in initial public offerings (IPO) of stock. In 1995-96, common law countries average 2.2 initial public offerings per million people while countries belonging to the French civil law tradition averaged 0.2 of an IPO, to the German tradition, 0.12 of an IPO, and Scandinavian countries 2.1 IPOs. Specifically, during that year Germany had 7 IPOs, France 10, the United States 803, and India 1114.[31]

These findings suggest that firms in common law countries tend to raise far more equity finance from public investors than do firms in civil law countries. They are especially striking because they hold across countries with dramatic differences in per capita GDP. For instance, India has more IPOs per million of population than any country following the French or German civil law tradition.

Moving to debt markets, countries from a German civil law tradition tend to have deeper debt markets than common law countries (on average, the private sector debt to GDP ratio is 0.97 in 1994). However, common law countries have larger debt markets (debt to GDP ratio of 0.68) than countries from a French (debt to GDP ratio of 0.45) or Scandinavian civil law tradition (debt to GDP ratio of 0.57).[32] Some of the differences in debt ratios may be because what is regarded as debt differs across countries. For example, unlike in the United States, fully funded pension liabilities show up as debt in German balance sheets which gives the impression that German companies are highly indebted when, in reality, they are not. Moreover, more of the debt is borrowed from financial intermediaries in Germany compared to the United States where public corporate debt markets are quite large. Therefore, it may well be that countries from a common law tradition typically have both deeper public equity and debt markets.

A final piece of evidence; accounting standards have been used in other studies as a good proxy for financial development.[33] Unlike measures of financial development based on market size, this measure of financial development is useful because it is not mechanically correlated with the amount of finance raised – which may be a function of the kinds of industries present in a country. Common Law countries again seem to be more financially developed than all but the Scandinavian countries because their rating on accounting standards is uniformly higher (on

[31] La Porta, et al. (1997, p.1137).
[32] La Porta, et al. (1997, p. 1138).
[33] See Rajan and Zingales (1998).

average, 70) than French origin (average 51) and German origin (average 63) civil law countries.

In summary, these authors find separately that Common Law countries protect investors better and financial markets are better developed in these countries. The natural conclusion is that these are cause and effect. But the theme of their work, ”Law and Finance” hints at a broader idea, that it is the legal system more than any other difference between the countries that leads to the differences in financial development.

4.3 Measurement Ahead of Theory

Even before looking at some new evidence, there are theoretical reasons to be cautious about drawing any conclusions, natural as they may seem. One has to make strong assumptions even to go from one finding to the other. For instance, while it may seem self-evident that more investor protection would lead to deeper capital markets in Russia – Lukoil trades at a fraction of the multiple of reserves that Western oil firms trade at because investors fear expropriation – it is by no means true that this principle extends to more developed economies. In fact, in earlier work we show that borrowing by large corporations is the least in the two G-7 countries, Germany and the U.K., which have the strongest creditor rights.[34] The reason may simply be that stronger creditor protection leads to inflexibility and potentially higher bankruptcy costs. This could lead to lower effort and investment incentives for managers. Therefore, the objective of a sophisticated legal system is not necessarily that of extending maximal and unconditional protection to all investors, but to ensure that various stakeholders' cash flow and control rights in the firm vary in such a contingent way as to maximize total output. This suggests a more nuanced line of causality from protection to financial development than that investigated by La Porta et al.

Equally worrisome is the question of whether these studies capture the true degree of protection enjoyed by investors in different countries. This cannot be understood without being able to specify the nature of the investor and the threat to investor rights.

La Porta et al. focus mainly on the protection granted to minority shareholders and creditors against managers and dominant shareholders. But civil law countries may want to protect a different type of investor than minority shareholders or creditors. For example, investing may largely be done through institutions in some civil law countries. This is

[34] See Rajan and Zingales (1995).

also consistent with the finding in La Porta et al. that the three largest owners hold a greater fraction of shares in French civil law countries than in common law countries. These investors may not need the protection afforded minority investors that La Porta et al. focus on. There is clearly a problem of causality that cannot be disentangled. If civil law countries do indeed have poor minority investor protection, is it because corporations there do not need to attract such investors since finance is largely intermediated by financial institutions, or is the direction of causality as La Porta et al. suggest; French civil law countries have few minority investors because minority investor protection is so weak.

Most importantly, none of the rights investigated by La Porta et al. appears unique to a common law system – the rights are present in at least some of the civil law countries. If there is one thing that the civil-law tradition is good at, it is writing new laws in great detail. Why then do we have this apparent contradiction with civil law countries not introducing seemingly simple, but valuable, statutes that seem to exist in common law countries? Perhaps the failure to introduce laws to protect investors is symptomatic of other, deeper underlying factors.

One potential piece of evidence that investor protection may be a symptom and not an underlying cause for financial development is provided by La Porta et al. themselves. When their measures of equity market development (such as the number of listed firms per million population) are regressed against the country of origin of the legal system and the index of antidirector rights, the latter is either not significant or marginally statistically significant. When the debt to GNP ratio is similarly regressed, the index of creditor rights has no additional impact once the country of origin of the legal system is accounted for. One possibility is that the aspects of investor protection that are most important for financial development are best proxied for by legal origin. Alternatively, there is something else that is associated with legal origin that leads to financial development but is unrelated to investor protection. We will examine both possibilities in what follows.

To summarize, the seminal work by La Porta et al. strongly indicates that countries who have a legal system rooted in English common law tend to have better developed corporate capital markets. More controversial is their claim that the reason for this finding is the higher degree of investor protection granted by Common Law countries, especially because the theoretical link between the two is still missing.

4.4 Common Law as a Process

A third possibility is that Common Law is more helpful than Civil Law in making operational legal devices such as ownership, contracts, and corporate structures, and social constructs such as good faith. Therefore, it may not be the specific statutes but the process by which the legal decision is arrived at that puts Common Law countries at the vanguard of financial innovation and development.

The focus of a common law system is on adopting the right process to reach a just solution. The civil law system, instead, expects the judge to apply the law as it is written. These differences would be irrelevant in a world where all future contingencies are anticipated and written in the initial contract or in the law. But contracts are typically incomplete in that they do not specify the rights and duties of all parties in all possible situations that may arise.[35] For example, when someone rents a house it is expected that he will subject it to reasonable wear and tear. What is meant by ”reasonable” is usually left unspecified, often because it is too costly to detail the myriad ways in which the house is abused. Instead, the owner and the renter negotiate after the rental period about whether the wear and tear is reasonable, and who will pay for it if not. If they do not agree, a court steps in.

Laws themselves are incomplete. While the judge may be equally capable of deciding what is just in a simple rental dispute in both the common law and civil law system, matters can be quite different as the legal form being fought over changes. We will argue that the focus of the common law system on adopting the right process to reach a just solution rather than creating the precise law up front allows it to deal with a far richer set of contingencies than the civil law system. This flexibility of the common law system allows more experimentation. Eventually, the case law stemming from such experimentation affords agents more certainty about the outcomes when a particular contract or form is used, and also results in a richer variety of contractual possibilities and legal forms. By facilitating contractual and legal adaptation, common law enables the ”production” of explicit and richly detailed arm's length contracts. This may partially be a factor why financial markets, entailing arm's length contracts between firms and investors, are more developed in common law countries. We now elaborate on this point.

[35] See Grossman and Hart (1986).

4.5 Legal Incompleteness – Some examples

It may be useful to start with some examples of legal constructs that differ across the two systems and to understand what might drive the difference. We look at two important constructs – the concept of ownership and the notion of the corporation.

The legal system is important because it defines the content of the ownership or property right and its limitations. The content refers to the object that is owned, while the limitations are the restrictions imposed on the ability to contract over that particular object. Traditional Roman law defined the ownership of a piece of land as an unlimited right *usque ad inferes et usque ad sidera (*from the center of the earth to the stars). More operationally, the owner has three prerogatives over the property in civil law – the right to usus (use), fructus (enjoy the fruits of), and abusus (sell or destroy). An important difference between civil law and common law, according to legal theorists David and Brierley, is that civil law does not allow easy fragmentation of these rights, while common law does.[36] This means that a variety of ownership structures are possible under common law, which are not under civil law.

For example, the concept of trust where the trustee has ownership rights and manages property in the interests of the trust beneficiary is alien to civil law. The trustee under Common Law has the right to administer and sell the property but does not use it and cannot destroy it. The beneficiary enjoys the fruits of the property, and perhaps the use, but does not have administrative control. Why cannot this structure which involves the fragmentation of ownership not appear easily in civil law countries?

The likely reason is that civil law cannot handle easily the myriad contingencies and exceptions that may arise in fragmenting ownership. What if the beneficiary and the trustee dispute a particular action such as the sale of property? Clearly, the resolution under common law would depend on the facts of the case, with the judgment trading off the need to be equitable to the beneficiary without hampering the administrative powers of the trustee. It is far harder under civil law to determine, a priori, where the balance of ownership rights should lie. Since legislature cannot make the determination, it is easier, though perhaps less efficient, to allow only simpler structures where ownership lies with the beneficiary and the trustee is simply the agent.

[36] David and Brierley , p. 324-325.

4.6 Corporate Opportunity

The ability to demarcate ownership precisely is not just important for relationships between owners and investors, it is also important in the creation and preservation of value. The value of a company is a combination of the value of its assets in place – such as plant and machinery – and the value of its investment opportunities – its plans for future growth. Assets in place are easily identifiable, and are protected by standard property law. By contrast, investment opportunities are harder to define and, as such, to protect. While it is fairly easy to prohibit an employee or director from competing with the corporation or from using corporate assets for personal use, it is not as easy to prevent her from taking advantage of a business opportunity.

Consider for instance the case of Guth vs. Loft. Guth, the president of Loft Inc a manufacturer and seller of beverages, bought the formula and trademark of the then bankrupt National Pepsi-Cola on his own personal account. He then built them into valuable properties, in part by having Loft buy lots of Pepsi Cola. The formula and the trademark did not belong to Loft. But the court ruled that the opportunity to buy and develop them did, and the firm had been unfairly deprived of the opportunity by its President. The court then forced Guth to turn the ownership of Pepsi to Loft. The rationale behind this decision was that Guth unfairly took advantage of a corporate opportunity that was rightfully the corporation's because it was an activity ”as to which it [the company] has fundamental knowledge, practical experience and ability to pursue, which … is consonant with its reasonable needs and aspirations for expansion” (Guth vs. Loft).

In the United States, the firm's right to opportunities is not protected by a specific law, but by a doctrine (the corporate opportunity doctrine). The concept is ambiguous. Courts disagree about the proper test for determining whether an opportunity is a ”corporate opportunity”, and case law attempts to define the doctrine more precisely. Nevertheless, the existence of this doctrine in the United States grants to an outside owner of the firm some control over investment opportunities, and allows her to capture at least some of their value. This makes possible, in the United States and in other common law countries, the financing of enterprises with valuable future opportunities but little assets.

The problems encountered in the United States in elaborating a universally accepted test of what constitutes a corporate opportunity illustrate the horrendous difficulties that would be faced in codifying this concept. Civil law would have to resolve much of the ambiguity up front. But ex ante complete codes are virtually impossible to write since many eventualities are impossible to foresee. Moreover, ambiguous

concepts such as intent and good faith play a part in determining the appropriate outcomes. These concepts are again highly contextual, and are more amenable to being judged ex post, based on fact, rather than ex ante based on theory. Of course, we do not intend to imply that the judge in a civil law system has no discretion, and does not use the facts of the case to arrive at a judgement. All that we imply is the empirical emphasis of the common law system may make it superior in both defining the property that is owned and devising an appropriate sharing of the fruits from, and the control rights over, the property.

A country with a civil code has two options in attempting to delineate property rights among different claimants to the firm better. First, it can try to describe rights in a variety of contingencies that can be anticipated, perhaps leading to an inflexible code. This may imply that certain statutes that exist in common law countries – because they can be tempered in operation by the interpretation of the courts – will not exist in civil law countries because they will be overly rigid in operation.

Alternatively, civil law can describe some broad-based rights that will suffice for most situations. The judge is then left to interpret in cases that cannot be easily pigeon-holed, with little guidance except universal principles of justice. In practice, ambiguity will not be resolved over time by the law since precedent does not play an important part in civil law.

4.7 Adaptation in Civil Law

We have probably overstated the differences between a civil code and a common law system. It is only a matter of time before the most important contingencies associated with a certain contract are observed and appropriate law written. Why cannot civil law replicate what common law does, but with a lag? Instead of the courts making law through cases, the civil code could be modified after a number of judgements reveal the difficulties with the existing law. After all, because the authority to modify the law is centralized in the legislature, a civil law system permits easy modification.

Adaptation may not be perfect. If the law is not crafted in the necessary detail to begin with, the courts may not be confronted with the necessary cases that would suggest a change in the law. For instance, if there is no corporate opportunity doctrine to begin with, it is hard for courts to confront the issue and build up experience that would suggest how the law should be rewritten to define more precisely corporate opportunities that are the firms. On the other hand, having a blanket opportunity doctrine without allowing courts to interpret which ones are

legitimately the firm's may also not result in the necessary experience. Few employees would break away and start their own venture for fear that the product of their effort, if successful, would be deemed to belong to the firm. This could stifle innovation and entry and, equally important, prevent the challenges to the law that lead to its development. In other words, the ability of courts under common law to frame and interpret the law also encourages cases that test and challenge the law and lead to its development.

Civil Law countries could imitate Common Law countries. However, the wealth of collective experience from case law is hard to codify into precise statutes. It is, in a sense, tacit knowledge that is common knowledge among the participants in the system, but extremely hard to express verbally. Thus while in theory civil law could evolve to capture the richness of the collective experience reflected in common law, in practice adaptation may be delayed and imperfect.

Nevertheless, as we will show, complicated legal constructs such as the ability to freely incorporate with limited liability have been quickly adopted in their final form by Civil Law countries, even though they were perfected after a long period of experimentation in Common Law countries. Thus it seems implausible that the inability to deal with legal incompleteness is what holds Civil Law countries back in the long run.

5 The Superiority of Common Law: A Historical Test

La Porta et al. present overwhelmingly evidence that Common Law countries are more financially developed. Their proposed causal link, however, is fragile. If financial development requires statutes that protect investors and these statutes are not unique to Common Law, why others countries do not adopt them? Thus, it must be the case that the advantage of Common Law is not replicable, like the *quality* of its enforcement process rather than its efficiency. At this point, however, this is only a conjecture, based on the current level of financial development. An independent test of this conjecture can be obtained going back in history. Since the choice of the legal system was done, at the latest, during the 19th century, we could try and look at the major indicators of financial development at the beginning of the 20th century and see whether it is still true that Common Law countries are more financially developed.

Answering this question is complicated by data availability. We have data only on a subset of countries in Table 1, and on only a few measures. To triangulate the ”true” measures, we present data coming from different sources.

The first source is Goldsmith (1969), who is the most authoritative source on the history of financial development. He reports some basic statistics for the most developed countries of the time. For comparability, we rearranged them in Table 2.

Column I of Table 2a reports per capita Gross National Product (GNP) in 1913 and column II the ratio of market values of securities outstanding in 1913 to GNP that year. In the definition of security Goldsmith includes both stocks and bonds, without differentiating between bonds of financial institutions and other bonds. This column, then, should be compared with the sum of columns (II) and (III) in Table 1.

The difference in data sources prevents us from drawing any strong conclusion about the difference in levels between 1913 and 1996. Nevertheless, it striking that all the values in 1913 are well above the corresponding 1996 levels. The closest figures are for the United States (1.5 in 1913 vs. 1.14 in 1996), the most dispersed for France (2.5 in 1913 and 0.5 in 1996).

Table 2. Financial Development before WWI

A: Period 1861-1913

	(I) Per capita GNP in 1913 $	(II) Securities/ GNP in 1913	(III) Securities/ Total GNP	(IV) Adjusted Securities/ Total GNP	(V) D Asset Banks Total GNP	(VI) (Securities + D Asset Banks)/ Total GNP
USA	406	1.504	0.069	0.053	0.040	0.093
Britain	313	2.268	0.071	0.029	0.026	0.055
France	232	2.454	0.074	0.040	0.030	0.070
Germany	211	1.533	0.063	0.030	0.061	0.091
Austria-Hungary	NA	NA	0.026	0.020	0.025	0.045
Italy	143	0.686	0.023	0.020	0.025	0.045
Russia	83	0.592	0.024	0.020	0.013	0.032
Japan	45	1.308	0.067	0.062	0.022	0.084

B: Period 1901-1913

	(I) Securities issued/ Total GNP	(II) Domestic sec. issued/ Total GNP	(III) Shares issued/ Total GNP	(IV) Bond issued/ Total GNP	(V) (Dom. Sec + D Asset Banks)/ Total GNP
USA	0.052	0.051	0.017	0.034	0.109
Britain	0.068	0.021		0.021	0.050
France	0.072	0.037	0.022	0.015	0.073
Germany	0.044	0.036	0.010	0.026	0.121
Japan	0.085	0.085	0.025	0.060	0.140

Source: Goldsmith (1969).

We can, however, be more confident about the differences within each period. In 1913 France and England appear the most developed capital markets, with Germany, Japan, and the United States close behind. Italy and Russia are a distant last, in part because of their relative economic underdevelopment. Thus, the initial first mover advantage enjoyed by England seems to have been eliminated. Although the paucity of observations prevents any statistical testing, there seems to be no connection between type of legal system and level of development. In fact, countries with a similar level of economic development (the first four) have similar levels of financial development. The variability within this group is much less pronounced than in 1996. The real outlier appears to be Japan. In spite of its late industrialization, it has a level of financial development comparable to the first four countries.

Unfortunately, for this time period it is difficult to have data on the level of capital formation financed with securities. Since most of the

securities outstanding in 1913 were probably issued in the period 1861-1913, Goldsmith approximates the securities issued during the period with the securities outstanding at the end of the period. He then normalizes them by the total GNP produced during the period (column III). Since Gross Capital Formation is roughly 20 % of GNP, to make these figures comparable to the one in Table 1 we have to divide them by 0.2 (i.e. to multiply them by 5). After that they should be compared with the last column of Table 1. The 1913 figures appear smaller than the 1996 ones for the United States and the United Kingdom, but much bigger for France and Italy. Since differences in the variable definition can explain more easily an overall difference in the level, not a cross-sectional variation in the difference, this pattern is likely to illustrate a real phenomenon: the importance of securities has grown in some countries, while it has declined in others.

Goldsmith tries to refine his estimate of the amount of corporate securities issued by eliminating bonds of mortgage banks, foreign securities, and an estimate of capital gains from the total value of securities outstanding. The ”adjusted” ratio is reported in column IV. The importance of security issues drops substantially for France and England, which had a large volume of foreign securities issuance (an issue we will discuss momentarily). The country that seems to have financed most with securities is Japan, followed closely by the United States, with the other developed countries close behind. Italy, Austria, and Russia are, again, far behind. Thus, the adjustment does not substantially modify the conclusions.

Column V we reports the ratio of the increase in bank assets to total GNP produced during the period. Excluding Russia, in all the other countries the increase in bank assets accounts for roughly 2.5 % of the GNP, except for the Unites States (4 %) and Germany (6 %). Thus, while banks appear very important in Germany, the traditional difference between bank-oriented countries and market-oriented countries does not seem to hold.

Following Goldsmith in column VI we sum the adjusted measure of securities issued and the increment in bank assets during the period. This measure is then deflated by the total GNP produced during the period. If the capital to GDP ratio is constant, the ratio in column IV provides a rough idea of what fraction of investment are financed externally. The difference between Anglo-American economies and the rest of the world, which emerged so clearly in Table I, is not present here.

Table 2a does not properly distinguish domestic securities from foreign securities and stocks from bonds. Goldsmith provides this information for a subset of countries and a shorter time period. We report this information in Table 2b.

Column I does not look substantially different from column II of Table 2a, in spite of the shorter time period and the different sources (these are actual data on securities issued rather than estimates). Column II reports the same ratio, where as numerator we have only dmestic securities. The difference between columns I and II suggests that London and Paris were the two main international financial centers, with Germany a distant third. Interestingly, the United States exhibits almost no foreign issues of securities.

In principle, the development of a financial system should be measured by its ability to provide cheap finance to industry, not by the actual amount provided. In fact, the actual amount provided is also a function of the demand of external funds by domestic firms. The fact that London and Paris (and to a lesser extent, Berlin) were able to provide more financing than their firms needed is an indication of their advanced level of financial development.

Column III reports the ratio of shares issued divided by the total GNP produced during the period. Japan and France rely most heavily on equity, followed by the United States and then Germany. Again there is no evidence that the equity market is more important in Common Law countries. The difference between columns II and III (reported as column IV) gives us an indication of the importance of the bond market vis-à-vis the stock market. Its importance is highest in Japan (6 %) followed by the United States (3.4 %) and Germany (2.6 %).

Finally, column V reports the ratio of the sum of domestic securities issued plus increase of bank assets to GNP produced during the period. Here the ranking seems to be negatively correlated with the period since the beginning of the industrialization process. The United Kingdom, which started first, finances very little of its total production externally – a fact consistent with Gerschenkron (1961). France, United States, Germany, and (at a distance) Japan follow. This result is not necessarily puzzling. The countries that started the industrialization process earlier had more established firms by the beginning of the 20th century, which could finance growth internally.

Goldsmith's data are gathered from a variety of sources. To validate his main conclusions we followed two alternative routes. First, the League of Nations *International Statistical Yearbook* has data on equity issues for a small subset of countries starting back in 1913. In Table 3 we report these data deflated by the GDP of that year.[37] For ease of comparison we report the same ratio computed in 1996 (data are from the FIBV).

[37] Data on GDP are from B.R. Mitchell, *International Historical Statistics*, Stockton Press.

Table 3. Equity issues over GDP

	1913	1996
Germany	0.014	0.004
Denmark	0.006	0.006
United States	0.047	0.029
France	0.053	0.015
Italy	0.013	0.005
Japan	0.118	0.002
Netherlands	0.056	0.053
UK	0.017	0.017
Sweden	0.055	0.009
Switzerland	0.133	0.007

Sources: League of Nations International Statistical Yearbook, Mitchell, International Historical Statistics, and FIBV.

Before interpreting the data, three caveats are necessary. First, the 1913 datum for the United States is in fact from 1923 (earliest available). Second, the 1913 reported datum for the United Kingdom represents domestic issues only, foreign subscriptions are five times as large. Finally, the 1913 data represent the actual issue value only for the United States, the United Kingdom, Switzerland and Germany. For the other countries they represent the nominal value, which is a lower bound for the issue value. Thus, they would underestimate the true value.

Switzerland and Japan appear as the biggest equity issuers, roughly 12 % of GDP. France, United States, Sweden, and the Netherlands are in second position with roughly 5 % of GDP. Germany, Italy, and United Kingdom are last with roughly 1.5 %. These results seem to confirm the one from Goldsmith reported in column III of Table 2.B. Again, it is very difficult to see a clear difference between Common Law countries and Civil Law countries.

Since the data, albeit collected from different sources, are fairly homogeneous, we can also attempt an intertemporal comparison. The importance of equity issues seems to have dropped dramatically in Japan, Switzerland, France, Sweden, Germany, and Italy. It has remained constant in the United Kingdom, Denmark, and the Netherlands. It has dropped a little in the United States. It is possible that this might reflect only seasonal factors. In Japan, for instance, equity issues have been curtailed in recent years, following a severe drop in the index. If we exclude Japan, the major drop is concentrated in Continental Europe. Is this just a coincidence or a more generalized phenomenon?

To explore this issue we make use of a third source. We obtained historical stock market capitalization data from *Global Financial Data*, a company that specializes in collecting historical financial information. It claims to be ”the most extensive, comprehensive collection of financial historical data in the world”. Yet, their coverage of market capitalizations around the world is quite sparse in the earlier part of this century. Data for the United Kingdom starts in 1900, for the United States in 1924, for the major European countries in 1929, for Japan in 1930.[38] In addition, we purchased data on stock returns from them, which are much more complete and start from the 1910s. Thus, we can compute the ratio of stock market capitalization to GDP for several countries in 1929. In order to obtain a measure of stock market capitalization for 1913 we extrapolated the data backward for a number of countries in two ways. The first simply takes the earliest datum available and extrapolates it backwards to 1913 using the return of the local stock market index. This estimate is correct if equity issues equal to delisting at every single point in time.

There are biases. Delistings (other than mergers which would have no effect for our purposes) probably take place at very low values of a stock’s capitalization, so they would have relatively minor effect. Issuances take place at high stock market values. So a more conservative estimate is to subtract the League of Nation data on equity issues from the final value of the stock market capitalization before extrapolating the data back using the actual return on the index. If equity issues were typically made when the market was high, we understate capitalization in 1913.

The League of Nation data start for most countries in 1913, but it does not report the information for the years 1914-1922. These are the years surrounding WWI where we do not expect many issues.[39] These data are reported in Table 4.

[38] Interestingly, for a number of Continental European countries data on stock market capitalization are available before WWII, but not after it until the 1970s. This, by itself, is a good indicator of the drop in importance of the equity market in these economies.

[39] Because the data on the market capitalization of the Curb Exchange (later Amex) starts only in 1938, in previous years we inflated the value of the NYSE capitalization by the ratio of the value of the Curb and the value of the NYSE in 1938.

Table 4. Evolution of Stock Market Capitalization over GDP

	1913 (a)	1913 (b)	1929	1938	1948	1960	1997
USA	0.68	0.50	0.75	0.64	0.31	0.66	1.33
Belgium	0.79	0.06	0.43	0.60	0.27	0.37	0.57
Bulgaria			0.05	0.09	0.00	0.00	
Canada	1.22		1.11	0.98	0.47	1.39	0.92
Czech Republic			0.11	0.15			
Denmark	0.92	0.90	0.38	0.28			0.55
Finland	0.37		0.34	0.32			0.61
France	0.66	0.27	0.43	0.22	0.12	0.27	0.49
Germany	1.93	1.34	0.29	0.17			0.39
Hungary	0.43		0.22	0.22	0.00	0.00	
Italy	1.71	0.42	0.32	0.29			0.30
Japan	2.56	1.82	0.79	0.99	0.05	0.27	0.53
Norway	1.59		0.22	0.20			0.43
Spain	0.53		0.33	0.42			0.55
Switzerland	1.75	0.97	0.63	0.63			2.26
United Kingdom	4.29	4.29	7.04	3.01	0.91	0.88	1.55

(a) assuming equity issues= delistings

(b) assuming equity issues = data from league of nations and delistings = 0

Sources: Global Financial Data, League of Nations International Statistical Yearbook, Goldsmith (for GDP data).

A few facts are worth noticing. First, the enormous value of the ratio in London. Note that this is not an estimate, but the actual datum. In the U.K. the magnitude of the ratio between stock market capitalization and GDP peaked in 1929, to drop dramatically in later decades. In fact, in 1960 this ratio was only 88 %. It would be useful to keep in mind that this was the period of nationalization in Britain.

Once we take out the United Kingdom, there is no difference between the rest of the countries. In fact, the relative size of the stock market in many European countries exceeds that of the United States.

Although the coverage is not continuous, in Table 4 we can follow the behavior of this indicator through the different decades of this century. Interestingly, all the countries experienced a major drop in the importance of the equity market after WWII. This is not true only of the defeated countries, but also of the victors. Only in the United States did the stock market regain its pre-war level of importance quickly. In all the other cases, the ratio in 1997 was below the ratio in 1913.

The overall results are striking. So striking that one worries they might be a statistical artifact. To eliminate this suspicion it would be useful to have some microeconomic evidence that the effect macro statistics are capturing is indeed real. Such evidence is provided by

Calomiris (1994). He compares the underwriting costs of new equity issues in Germany and in the United States at the beginning of the century. Underwriting costs can be considered a proxy for the efficiency of capital markets (the smaller the cost, the higher the efficiency). The spread charged by U.S. banks on issues of common equity was about 18 %, while the one charged by German banks was only 5 %. This would also explain why between 1896 and 1913 German firms issues substantially more equity than U.S. firms.

Summary

We started this work noticing the variability of various indicators of financial development across fairly homogenous countries and we asked why such differences existed. We can rule out at least one answer, with all the caveats about not having enough data to perform tests of statistical significance. Differences in the nature of the legal system cannot be directly responsible. The legal system did not change over the decades, so why did the level of financial development? To explain why France and Germany are less financially developed than the United States today, we do not have to look at the distant past. We need to understand what happened in between the two wars, that changed the course of financial history. Having ruled out the other major causes, we are left with only one: politics. This is what we explore next.

6 A Political Economy of Financial Development

Given all the benefits of financial development, it is somewhat surprising how few countries in the world have sophisticated financial markets. Clearly, a well developed legal, accounting, and regulatory infrastructure are necessary for a well functioning financial system. Why don't countries copy more successful models? Fixed costs, while initially important, are not a major impediment in the long run nor, as we have argued, is the nature of the legal system. The impediment, we believe, is to be found through an analysis of the political economy of financial development. In this section, we attempt a simple calculus of consensus à la Buchanan.

The two main players in this analysis are the interest groups and the Government. Interest groups pressure the government to favor them. The nature of the government – whether democratic, oligarchic or a monarchy – will determine which group has more saying and to what extent its wishes get translated into actions. Until the beginning of the 20th centuries governments were at best oligarchic. Thus, the calculus of consensus can be limited to the dominant elite, which had the biggest impact on the government. In the 20th century the analysis is further complicated by the growing influence of the masses.

But the government does not simply reflect the wishes of the dominant interest group, it also has its own agenda. This is clearly the case in a monarchy, but also in a democracy where the entrenched government bureaucracy has its own perks and survival at stake. Thus, asking why some political systems oppose financial development is tantamount to asking why some influential groups or the Government (or both) oppose financial development.

6.1 Why the Dominant Elite May Oppose Financial Development?

To illustrate in a simple way the various political forces at work, we assume that there exists a magic law able to eliminate all the imperfections present in a real world financial market, transforming it into a textbook-like perfect capital market. This fiction will allow us to

study, in a highly stylized framework, who will gain and who will lose by an improvement in financial markets and, thus, identify supporters and opponents of financial development. It will also identify when this opposition is more likely to be successful and what forms it might take.

One of the main consequences of the approval of our hypothetical law will be a reduction in the rents from the ownership of current resources; In an imperfect world , some own-capital is needed to fund projects – outsiders do not have the information or control to fund a hundred percent. By contrast, in a perfect capital market, projects can be funded entirely from financiers, and the proverbial inventor in a garage can bring his ideas to fruition without possessing any family wealth.

Thus, the approval of our hypothetical law will have two main distributional effects. First, people with an initial endowment of resources, whom we will call the rich (or equally, the well-connected) will no longer earn an excess return because only they can get opportunities financed. Second, competition in output markets will increase because of increased entry from those wealthy in ideas, diminishing the rents existing producers might have enjoyed. Since existing producers tend to be owned by today's rich, the rich will lose again.

Another distributional consequence of the approval of our hypothetical magic law would be the loss in value of reputation. In a perfect capital market, i.e. a market with perfect disclosure and perfect enforcement, reputation is worthless. By contrast, in a world with no disclosure and no enforcement, investors will have to rely exclusively on the reputation of the issuer. Thus reputation, very much like the initial endowment, becomes necessary to start any economic activity. But reputation is something that takes time to accumulate, cannot be purchased, and can be transferred to others only with difficulty. Since the rich or well-connected typically are better known, they will lose from an increase in transparency and enforcement.

Capital markets also tend to point out inefficiencies and cross-subsidies, and increase the risk that underperformance will be penalized. Moreover, they are brutal in that they do not recognize past investment or favors unless explicitly contracted for. If contracting is imperfect, those who are likely to underperform, or legitimately fear the risk that they might underperform, will lose from financial development.

Of course, a perfect capital market increases the availability of funds for profitable projects and it improves the return on capital invested. These benefits occur to everybody. Thus, the extent of opposition depends on whether the share of benefits a particular group can appropriate is sufficiently large to compensate for the losses suffered.

For example, if the rich have become so because they had ideas in the past, and will continue to have them in the future, the benefit produced by an increase in the number of projects they can undertake will more than compensate for the reduction in the value of their initial endowment. By contrast, if the rich made their money in the past through passive ownership of resources such as land, the loss they suffer in the value of their endowment and reputation may dominate any gains. In other words, an economy dominated by self made entrepreneurs is likely to push more for financial development than an economy dominated by the landed gentry.

6.2 Financial Markets and the Government

The development of financial markets affects the government's own agenda. On the one hand, it makes it easier for the government to raise financing. On the other, it creates private sector competition for the funds of citizens and foreigners, which may reduce the government's share or increase its cost of funding. Some governments may find it better to keep the market underdeveloped so that citizens have no option but to park their funds with it (or its agents).

Of course, the system of government affects its willingness and ability to intervene. A centralized power is the biggest impediment to financial development. When the property rights of citizens are not well developed, strong governments cannot commit not to expropriate. Almost by definition, financial markets and institutions tend to concentrate a lot of wealth in one place. In so doing they make it easy for the Government to grab resources at a very low cost (or, in more modern terms, to collect taxes at a low cost).[40] Of course, the Government incentive to expropriate the financiers depends crucially on who the financiers are and who controls the Government. A monarch supported by the landed aristocracy typically found irresistible the temptation to expropriate the financiers from the city, especially the foreign Lombards and the Jews. By contrast, law bound (and weaker) [41]Governments cannot expropriate, allowing finance to flourish.

Even when naked expropriation is not possible, one cannot escape the fact that powerful centralized governments usually have agendas. Free

[40] See Myers and Rajan (1998).

[41] This shows the power of anti-usury laws in preventing financial development. Not only did they prevent a large fraction of the population from entering the financial markets, they also made it easier to the monarch to expropriate the financiers.

markets typically have only one agenda, to make money. It is always tempting for the government to alter the rules of the market so as to direct the resources of the market in a favorable direction, rather than pay the appropriate price. This is especially true when the government faces an outside threat to its survival – when it wants to transform the entire economy into a command and control system because that, rather than uncoordinated markets is the quickest way to mobilize resources. The long term damage to markets can, of course, be considerable. For example, centralized power has an incentive to create oligopolistic banking systems to replace financial markets, because these can be persuaded with favors and threats unlike an arm's length market. Moreover, once created, hierarchies are concentrations of power, and unlike dispersed markets, can coordinate to oppose change.

Markets also require and create a high degree of transparency, which can highlight the inefficiency of the actions of centralized power. This is inconvenient, especially in a democracy (democracies, of course, can be centralized).

Finally, centralized power might undertake activities that would otherwise be undertaken in the private sector and further the development of markets. For example, a railway system became imperative for a country's infrastructure during the 19th century. As already mentioned, this spurred the rise of financial markets in the United States. In many Continental European countries, however, the fear the private sector would lag behind in railway construction lead to direct Government intervention. By directly financing the construction of railways, the Governments may have overcome the short term disability created by inadequate private financial markets, but unwittingly prevented the emergence of private institutions that would have found railway financing a good way to overcome fixed start-up costs, with negative long term consequences. Similarly, during the 16th and 17th centuries the power of the Spanish fleet, which protected the ships traveling back from the New World, eliminated much of the need for insurance This greatly retarded the development of the insurance markets in Spain as compared to England or the Netherlands.[42]

[42] Santos (1998).

6.3 Interaction Between Politics and the Law

Even if the nature of the interest groups that hold political power at the center or the Government own agenda are the primary determinants of financial development, the legal system can play an important role, by affecting the influence these groups have. In particular, there are several reasons why the Common Law system can better resist the political pressure against financial development.

First, the English common law tradition emphasizes restraining the government and protecting the individual, making more difficult for the Government to interfere with private property. Thus, in the early stages of financial development, Common Law helps protect the rights of the individual against the state, and hence fostered the development of private finance.

Second, Common Law evolves at the periphery, and innovates around legislative or administrative roadblocks set up by the center. In England, for instance, in 1720 an act of legislation (the Bubble Act) placed constraints on the incorporation of limited liability companies. Common law courts, however, made their own judgement as to which companies did not contravene the spirit of that law. Thus, the decentralized nature of decision making enables Common Law countries to overcome the anti-market dictates of the Government.

Finally, Common Law is based primarily on judicial precedent, rather than upon the act of a legislator. Thus, a new political climate is less able to transform the structure of the law. For example, Roosevelt's attempt to cartelize the U.S. economy failed not because of political opposition, but because of the resistance of the U.S. Supreme Court, which struck down several key pieces of the New Deal legislation. In Continental European and in Japan Governments did not find such resistance and successfully cartelized many sectors, starting from the banking one.

7 Back to History

In the previous section we concluded that there are two important factors determining the likelihood of financial development: the calculus of its costs and benefits for the different political constituencies and the power and centralization of Government. In what follows we undertake a brief historical analysis of the conditions surrounding the emergence of major financial centers to verify whether they are broadly consistent with the predictions we highlighted.

7.1 Early Financial Centers

Till the late middle ages, the dominant elite in many European countries were the landed aristocracy. Traditionally, farming was a very low-capital intensity business, but for the cost of the land itself. The aristocracy inherited the land or received it from the Crown, in return for which it met its feudal obligations. There was little to benefit from a financial system; Trade was considered degrading (itself evidence of a cultural bias against new comers), and most wealth was embodied in land, which generated little investible surplus. However, the aristocracy had a lot to lose from financial development. First, the emergence of a merchant class could jeopardize its political domination. Second, the ease of access to external funds in the absence of profitable investment opportunities could translate into a surge of luxury consumption by the young nobles eventually leading some aristocratic families losing their land. This would have lead to a separation between land ownership and nobility, which could undermine the stability in the ruling elite. The desire to prevent such disruption can explain why, in a society dominated by nobles, it was so difficult, if not impossible, for the nobles to sell their land.

Contrast this with Venice, the epitome of the Italian city-state. Venice was born as an insular power, with no territory on the mainland. As a result, the aristocracy who ruled the Republic were essentially mercantile, and could benefit greatly from financial development. Not surprisingly Venice emerged in the late Middle Ages as one of the financial capitals of the world, the cradle of major financial innovations. Double entry book-keeping was invented there by Fra Luca Pacioli, as were certain maritime insurance contracts. The Venetian commenda, a

form of limited partnership, is also the most direct ancestor of the modern joint stock company. Venetian Public Debt was bought by foreigners, drawn by Venice's reputation for servicing it through thick and thin – in fact, Venice's Monte Vecchio served as the equivalent of Swiss Banks of our time, holding funds with no questions asked.

It is also not coincidental that power in Venice was surprisingly widely distributed among the citizenry, considering the times. Venice was a republic with checks and balances on the power of its government. Its executive, legislature, and judiciary, right up to the Doge, were elected, and except for the Doge who was a life appointee, held tenure for only a short time. The short tenure not only protected against incipient dictatorship, it also ensured that all eligible citizens got a turn at governance. There was always a strong chance that the roles would be reversed in the near future when a supplicant faced an official. This implied that governance, at least for eligible citizens (approximately one fortieth of the entire population) was likely to be fair. As insurance, those that displayed dictatorial tendencies while in office were often banished.

Consistent with the political theory, all the early financial centers of the Modern world, starting with Florence and Venice, going on to Hamburg and the cities of the Hanseatic League, and culminating in Amsterdam, emerged in little independent political entities controlled by the bourgeoisie (bourgeois literally means inhabitant of the city). The bourgeois was not a landlord but a craftsman or a merchant, who needed access to external funds to finance his working capital. By contrast, none of the political centers of big and powerful centralized states (Paris, Madrid, or Vienna) emerged as important financial centers.

7.2 The Emergence of London

England was the first of the large nation states to have a well developed financial system, and London was at the center of it. How did London emerge as a financial center in the 17th and 18th centuries? In part, it was the devastation produced by the two Dutch-Spanish wars during the 17th century, which significantly weakened Amsterdam. But it was also the fortuitous coincidence of the two crucial catalysts predicted by the theory: an elite who could benefit from financial development and a weak Government. We review them in turn.

In England, as in most other Nation States at that time, the dominant elite was the landed aristocracy (gentry). But unlike other States, the English landed aristocracy in the 17th century could benefit from a development of the capital markets. The agricultural revolution, started

by the enclosure movement, transformed British agriculture into more of a business. And as a modern business, British agriculture required capital and generated a surplus that could be invested. Thus, British aristocracy, unlike aristocracy elsewhere, had valuable opportunities to invest extra funds and, after the initial ”dividends”, more funds to be reinvested in the capital market. This made it better disposed towards the development of a financial system. One should also not understate the strength of the mercantile constituency, fortified by England’s maritime prowess.

At the same time, financial development was facilitated by the weakness of the British Crown. The Glorious Revolution of 1688, which put William and Mary on the throne, also forced them to agree to a Declaration of Rights. By doing so, the Crown recognized the legislative supremacy of Parliament and also the necessity of parliamentary consent for a standing army in peacetime. Given that Parliament represented both the moneyed interests of merchants and financiers (the Whigs) and the landed gentry (the Tories), it was clear that property and financial contracts would be much more secure against the depredations of the Crown. Equally important, without a standing army, the Crown had no way of imposing its wishes on Parliament. Finally, the existence of an independent judiciary enforcing common law further strengthened the hands of public creditors against the Crown. As political economists Douglass North and Barry Weingast argue, the credible commitment not to interfere with property rights led to the development of government finance. So, somewhat paradoxically, as sociologist Bruce Carruthers argues, the internal weakness of the English Crown allowed it to raise large sums at short notice, giving it external strength and transforming England into a first rate European power.

It did, however, take time for the English Crown to gain credibility, even after the Glorious Revolution, especially because of previous events such as the Stop on the Exchequer, when Charles II suspended debt payments amounting to about $ 1.3 million pounds (at a time when annual Crown income was less than 2 million pounds).[43] Therefore, much of its initial funding was not direct but through large companies, prominent among which were the Bank of England, the East India Company, and the South Sea Company. By 1712, these three companies had loaned 15.8 million pounds to the government, which was approximately 62 percent of the funded debt.[44]

Each of these companies was granted a monopoly, which, in effect, was an indirect tax on the people. The Bank of England had a monopoly

[43] Carruthers, p. 122.
[44] Carruthers, p. 155.

over note issue, the East India company had a monopoly of trade with the Indies, and the South Sea Company obtained the monopoly over trade with the South Seas (a monopoly which turned out later to be of little value). In order to secure and preserve these monopolies, the companies repeatedly loaned money to the Crown. Thus the early joint stock company was set up, in part, as a device for the Crown to transform future monopoly rents into current funding. Since the shareholders of these companies were many, and included many prominent members of Parliament, the commitment made by the Crown (with the acquiescence of Parliament) to the companies was secure.

While shareholders of the companies knew that they were secure from government expropriation, it was less clear that the management of the companies could not fiddle with the accounts or otherwise exploit investors. The absence of corporate governance would surely have hurt the companies' ability to raise money from investors, and thus also hurt the government, which depended so much for funding from these companies. Therefore, it was in the interests of the government to ensure the public investor was protected. Moreover, the companies were so intertwined in politics that it was almost inevitable their governance would resemble the emerging democratic political process. This was especially so because the two strong competing political parties each wanted a share of the power and patronage that came with control over the companies.

As a result, early corporate charters, to which the Government had to give its assent, ensured that the directors of the Bank of England and the East India Company were elected. Even though voting was restricted to those who had a substantial stake – the East India Company allowed only those with over 500 pounds in shares to vote – the election of directors was a serious event, and trading at the time of the election of directors was active, and especially so in blocks over 500 pounds.[45]

The negative influence that a strong government can have on corporate governance can also be seen in some periods of British History. After unsuccessful attempts at gaining control through elections over the Bank of England and the East India Company in 1710-11, the Tory government of Robert Harley, which was backed by an overwhelming majority in Parliament, introduced a bill proposing the South Sea Company. Unlike the other two prominent companies, the power to appoint the first court of directors was vested in the Queen, with Harley becoming governor. Thirteen of the appointed directors were from the list of unsuccessful Tory candidates for the Bank and East India Company elections. Of course, Harley could not completely

[45] Carruthers, p. 151.

antagonize the Whig moneyed interests because they were the primary source of funding, so some investor protections were built in. The point, however, is to suggest that the initial fortuitous political climate in England, as well as the need for government finance, led eventually to the development of governance of joint stock companies.

7.3 The French Experience

As the political scientist and historian, Samuel Finer puts it, "the four great differences between the absolutism of the king of France [in the seventeenth century], and the relative limitation of the English monarchy were…that the former could strike taxes and make laws at pleasure and was served by a standing army and paid professional bureaucracy". There was a consistency in these powers. The King could levy new taxes, which the bureaucracy would collect, with the threat of the standing army always in the background to enforce payments. Taxes, though, take time to levy and collect, especially when the bureaucracy is as inefficient as the French then was. Thus the Crown had to borrow to finance wartime expenditure. A number of economists and historians have argued that the immense power of the French Crown, and its past record of defaults, prevented it from raising loans directly from the public. Furthermore, it was impossible to protect investor rights because of the unlimited personal authority of the king.

Instead, the French Crown raised money at short notice in two ways. First, it borrowed through lesser state bodies such as the Hotel de Ville de Paris. The advantage of borrowing through these lesser state bodies is that, unlike the king, they could be sued in court.[46] Of course, since the King ultimately also had power over the courts, this turned out to be a protection of doubtful value. Second, the sale of venal offices was another source of funding. The purchaser paid for the office up front, and made additional investments down the line to acquire additional powers. In return, the government paid an annual interest, and conferred status (occasionally, minor nobility) on the purchaser. In the 1660s, forty six thousand offices had been sold (with many offices being split over time as the need for new funding arose) for an amount approximating four to five times the annual revenue of the crown. Of course, there was no requirement that the purchaser be qualified for the

[46] More valuable was the right that if a city failed in its obligations to a creditor, the creditor could take actions against the person and property of any citizen of that city who fell in his power. Thus cities that traded a lot obtained the ability to borrow by giving creditors this extreme power.

office, and since many of these offices included actual duties such as collecting taxes or serving as magistrate, the efficiency of administration fell over time. Nevertheless, unlike loans to the crown, the offices carried their own security since the Crown could not do away with them with destroying a substantial part of the administrative machinery.

The French Revolution did not change matters much because, while power was taken away from the crown, it was concentrated again in the central government. The French did not have a true parliamentary democracy until the constitution of the Third Republic was passed in 1875.[47] It is useful to see the effects on two major components of a modern economy – limited liability corporations and the banking system. For reasons of space, our description will only be a sketch.

The legal forms of business organization permitted by the French Civil Code of 1807 were simple partnerships, limited partnerships in which only the active partners had unlimited liability, and true joint stock or limited liability companies. In order to incorporate as a limited liability company, a firm needed special permission from the Council of State. This body granted charters sparingly, and usually only to industries it deemed in the public interest.[48] It was not until the depression of 1857, and under pressure from the now strong industrial interests that the state was persuaded of the greater resilience of limited liability companies, and passed laws in 1863 and 1867 that allowed free incorporation.

Contrast this with the developments in England. While the Bubble Act of 1720 did place constraints on the incorporation of limited liability companies, common law courts made their own judgement as to which companies did not contravene the spirit of that law. So, for example, in 1811, the issue of widely held, and transferable, common stock by an unincorporated bread manufacturer was deemed to be legal.[49] Since allowing the transfer of a stock essentially implied that its previous owner could not be held liable for the firm's debts, this judgement allowed limited liability for a venture if its intent seemed honorable. Therefore, even though the Bubble Act was repealed only in 1825, and companies could get limited liability without the necessity of a charter from the government in 1855, common law allowed many of the trappings of limited liability to corporations much before then.

There are two important points to reiterate from this discussion. First, common law allowed more legal and financial innovation (for example, in corporate form). Second, the civil code could speedily adopt

[47] Finer, p. 1595-96.
[48] Cameron, p. 112.
[49] Powell, p. 178.

innovations that emerged under the common law system when it so desired. It took over a century and a half for the common law system to work out the limited liability form to its satisfaction in 1855, and at the risk of some understatement, only about a decade for the French civil code to adopt the main outlines of this legal innovation.

This then reinforces our claim that the greater ability to innovate is, by itself, not the reason the common law system fosters financial development. Instead, it is partly the ability of the decentralized common law system to innovate around the political impediments placed by the center that enable it to foster financial development. By contrast, in civil law countries, the centralized control of the legislative process ensures that the powers that be can stand in the way of anything against their interests.

To illustrate this point, consider the differences between the broad actions of the Bank of France and those of the Bank of England. After Napoleon came to power, he sought a ready source of credit. Given the success of the Bank of England in mobilizing credit for the English government, a bank seemed a natural device. The Bank of France was capitalized at 30 million francs, but unlike the Bank of England, had ”difficulty placing its shares in spite of the personal example of the First Consul and his decree requiring government agents to purchase shares and deposit their surplus funds in the Bank; almost two years elapsed before the entire capital was paid in”.[50]

The Bank of France was little more than an extension of the government (perhaps indicating why it had difficulty raising capital), and it had extensive demands made on it. With the fall of the Empire in 1814, the Bank went into virtual liquidation. While there were some moves to privatize the Bank at this point, the Restoration government felt it was too valuable a resource to lose. As Baron Louis, the then Finance minister told the representatives of the Bank ”You want to be independent, but you will not; you will have a governor, I will name him, and he will not be the one who currently occupies the post”.[51]

Between 1800 and 1848, the Bank of France stood as a bulwark against financial development because competition threatened its position.[52] The Bank opposed the setting up of joint stock banks. It initially resisted setting up its own branches in other cities, and did so only the cities started setting up competing establishments. But having determined to expand outside Paris, it quickly got the government to refuse to charter any new banks. It also took advantage of crisis to take

[50] Cameron, p. 102.
[51] Cameron, p. 103.
[52] Cameron, p. 104-106.

over competing note issuing banks (obviously with government support) so that in 1848, its monopoly over note issue extended over the whole of France. Interestingly, it was only after the Revolution of 1848 that financial development resumed. Again the reason was political. As economic historian, Rondo Cameron puts it, ”the characteristic institution of the Second Empire, the Societe Generale de Credit Mobilier, also came into existence in 1852 as a part of the government's attempt to provide a counterweight to the influence of the Bank of France and the financiers of the haute banque, most of whom were hostile or at best lukewarm to the new regime”. Finally, even though the monopoly of the Bank of France was slowly whittled down, it had lasting effects. As Cameron argues,

> ”this failure [to reorganize the commercial banking system], which affected even large enterprises…was especially serious for the numerous small industrialists who could not interest the great bankers in their enterprises or, if they could, feared loss of control and independence…Thus resulted the schizophrenic development of French industry and commerce, traces of which can still be seen today: on the one hand a highly capitalized, technically efficient and progressive sector, on the other a sector composed of small firms, undercapitalized, traditional but tenacious.”

The contrast with England is striking. While the Bank of England had similarly tried to protect its privileges against encroachment, it was far less successful because it was a private entity, because it was in the interests of the Tories to see its powers contested, and because the common law system allowed alternatives like the country banks to emerge. As a result, there was far more competition in financial services, with attendant benefits to industry.

7.4 British and Spanish colonies

The differing degree of centralization and the nature of the powerful elite can also explain why the financial development of British colonies today appears so much greater than those of former Spanish colonies. As historian Samuel Finer argues, the English were either escaping religious or political persecution, or trying to improve their lot. They immigrated in entire communities, with people from every social strata and skills represented. A goodly portion represented non-conformists, indentured servants, and even convicts. Thus, the homogeneous social composition of English colonies favored a political consensus for financial development, as did the entrepreneurial spirit of the early colonizers.

The English colonies also benefited from a weak and decentralized Government. Their political system reflected the tension between Crown and Parliament, leaning to the Parliamentary side. The colonies were certainly the Dominions of the Crown, and yet not one colony lacked a legislature whose minimalist function was to set taxes.[53] In fact, some colonies were essentially joint stock corporations, with the original Massachusetts charter providing that the Governor and his eighteen assistants had to be elected annually by the stockholders and furthermore, requiring quarterly meetings of the entire body. The English government essentially viewed the Empire as a huge trading organization, and interfered with the activities of the colonies only in so far as it affected trade.[54]

The Spanish colonies were quite different. The Spanish immigrants were "…for the most part, adventurers, soldiers and officials, in short, a ruling stratum".[55] Their purpose was largely extractive. As Adam Smith puts it [56]

> "It was the sacred thirst of gold that carried Oieda, Nicuessa, and Vasco Nugnes de Balboa [to America]. Whenever those adventurers arrived upon an unknown coast, their first enquiry was always if there was any gold to be found there; and according to the information they received concerning this particular, they determined either to quit the country or to settle in it."

The prospect of uncountable wealth implied that the Spanish Crown always had a strong interest in its colonies. Spanish colonies were organized as hierarchies directly under the control of Castile. All important decisions, as well as many unimportant ones, were made in Spain, and a stream of rules and laws flowed to the colonies with the purpose of binding and guiding the administration.

The Spanish, therefore, transported their hierarchical system and their aristocracy (at least the attitude and mentality of their aristocracy) to the colonies. The deficiencies of the centralized system that we have referred to earlier, again demonstrated itself. Adam Smith, with typical understatement, writes[57]

[53] Finer, p. 1383.
[54] Finer, p. 1399.
[55] Finer, p. 1382.
[56] Smith, Book IV, Chpt VII, p. 73.
[57] Smith, Book IV, Chpt VII, p. 79.

> ”The Spanish colonies, therefore, from the moment of their first establishment, attracted very much the attention of their mother country; while those of the other European nations were for a long time in a great measure neglected. The former did not, perhaps, thrive the better in consequence of this attention; nor the latter the worse in consequence of this neglect.”

And later,[58]

> ”There is more equality, therefore, among the English colonists than among the inhabitants of the mother country. Their manners are more republican… The absolute governments of Spain, Portugal, and France, on the contrary, take place in their colonies; and the discretionary powers which such governments commonly delegate to all their inferior officers are, on account of the great distance, naturally exercised there with more than ordinary violence.”

Thus the English colonies were set up in such a manner as to facilitate the political forces leading to financial development. By contrast, the colonies of Spain (and France) accentuated the forces that served to hinder financial development.

7.5 The Contrast with the Netherlands

An interesting, middle of the road, example is represented by the Netherlands. The Union of Utrecht in 1579 created a decentralized structure where the members of the States General, the national parliament, were selected by the different provinces. Delegates had to refer back to their respective provincial assemblies often for instructions.[59] Even though the Dutch was one of the most heavily taxed people, the United Provinces did not have a centralized tax system. Money was raised at the provincial level and then funneled to the center. Each province had a quota, and extraordinary wartime expenditures were appropriately shared. Large and sudden expenditures were financed through public borrowing.

The Dutch government, unlike the English Crown, established a reputation for servicing its debts quite early. In part, this was because the democratic tradition emerged earlier, allowing political curbs to be placed on the powers of the government. Since government debt was widely held (around 1670, the Dutch Mercury stated that in the province of Holland alone, there were 65,500 persons who had or were able to invest money in annuities), much of it by the ruling elite, there was little

[58] Smith, Book IV, Chpt VII, p. 98-99.
[59] Carruthers, p. 103.

chance that the government would willingly default on its debt.[60] In fact, ”investment in home securities was so popular that it was often considered a favor to be allowed to invest in them; and repayment was a source of regret and even called forth tearful remonstrances from creditors, because in no other place could money be put out so quickly and safely”.[61]

Contrast this reputation with that of the English Crown before the Glorious Revolution. It is no wonder that the Dutch government could borrow at an interest rate of 4 percent while Charles II had to pay up to 12 percent on his borrowing.[62] Even after the Glorious Revolution, the English Crown had to establish a reputation for repayment, and convince investors that their power to secure repayment was adequate, before it could borrow large long term funds directly from the public. As a result, it had to rely substantially on the joint-stock companies for funding. By contrast, the Dutch Republic, which could borrow directly, never felt the urge to ”convert public debt into company shares”. In fact, several provinces and cities like Amsterdam, Haarlem, and Leiden, forbade the establishment of new joint stock companies.[63]

So both in England and the Dutch Republic financial development could take off because the elite wanted it and power was decentralized. The major difference between the two countries was in the political strength of the two Governments. The English Government, caught in the struggle between the Crown and Parliament, was politically weak. The Dutch Republic, instead, could count on a strong popular base. Thus, it had no need/incentive to favor the formation of investor friendly corporations. A second difference was that even though the Dutch parliament was indirectly elected, it was firmly under the control of the provinces, and by extension, influential investors. This then seemed to be an adequate means of governing the Dutch East India Company (also known as the VOC), the largest and most famous of the Dutch joint stock firms. Operating control of the VOC was vested in 17 active directors, 8 of whom were nominated by Amsterdam, 8 by the other large provinces, and one directorship rotated among the smaller provinces. In short, since the state enjoyed the confidence of the public (at least, the influential among them), there seemed little need for a separate governance structure for corporations. Shares had no voting rights, and trading was especially active at dividend time, in contrast to the English common stock companies where trade was especially

[60] Carruthers, p. 104-106, Ehrenberg, 350-351.
[61] Ehrenberg, p. 350.
[62] Carruthers, p. 60.
[63] De Vries, p. 153.

important at the time of election of new directors.[64] Because of the greater credibility of Dutch public governance, independent corporate governance did not develop in the Dutch Republic to the extent that it did in England.

7.6 The Triumph of the Bourgeoisie

On the basis of the historical account so far it is hard to separate the influence of the legal system and that of politics. As Cameron puts it, ”Fortunately for England, the law was sufficiently loose and its administration sufficiently lax that the obstacles to innovation were not insurmountable”.[65] So was it the law or was it politics? Only the subsequent events of the 19th century provide us with a chance to disentangle the two effects. During the 19th century the opportunities provided by the Industrial revolution made it extremely costly not to have an efficient financial system. At the same time, various Revolutions brought the bourgeoisie in power in most countries in Europe. In fact, according to Marx (1848), governments became business committees of the bourgeoisie. Thus, by looking at the level of financial development of the most important developed countries at the turn of the century we can distinguish the two hypotheses. If the main explanatory variable is the legal system, then countries with a common law system should stand out for their level of financial development. By contrast, if politics is important, then all developed countries should have a similar level of financial development.

The evidence reviewed in Section 6 is supportive of the latter interpretation. Besides England, which, thanks to the early move, continued enjoying the benefits of being the financial capital of the world, all other developed countries had a similar level of financial development. Interestingly, many countries that today had underdeveloped markets, back then had very flourishing markets. Thus, a persistent variable, like the nature of the legal system, cannot by itself explain the retreat of the market. It can, however, if coupled with a change in the political environment. This is what we are going to explain next.

[64] Neale, p. 9.
[65] Cameron, p. 59.

8 The Reaction

World War I and its aftermath changed everything. As Table 4 shows, financial markets declined significantly in many countries between 1913 and 1948 even as a proportion of GDP, and in some cases have not reverted back to their prewar level even today. What happened?

That World War I was a great disequilibrating event is probably an understatement. Apart from changing the map of Europe, and destroying the people's faith that their rulers knew what they were doing, the long war accentuated the centralization of many of the economies of Continental Europe. McNeill (1981, p. 339) describes Germany after the implementation of the "Hindenberg Plan" of 1916 thus

> "…the generals in charge often became impatient with the financial claims and controversies that continually embroiled and sometimes obstructed prompt and deferential obedience to their demands. As shortages rose, one after another, the generals relied more and more on big labor and big business to remodel the economy according to military needs. Each party got more or less what it wanted: more munitions for the army, more profits for the industrialists, and consolidation of their authority over the work force for union officials."

In other words, the economy was being run like a centralized hierarchy model along the lines of the army. Centralization was aided by cartelization – of industry, of the banking sector, and of the work force – because this reduced the number of parties the central authority had to negotiate with. Prices were fixed, and distortions built up in other sectors, especially agriculture, which finally led to Germany's surrender.[66] But the aim of maximizing military production was served, and the costs papered over by blaming the weak-kneed politicians for the surrender. For countries like France, however, the benefits of command mobilization were further sweetened by victory.

The centralization during the war thus created large domestic hierarchies representing different interest groups, all of which had acquired power when shielded from international markets and competition. By contrast, the groups that would naturally support markets and competition were suppressed and weakened by the relative

[66] McNeill, p. 340-341.

autarky forced by the long war. The anti-market groups therefore had more powder, and it was of recent vintage.

Centralization also created large distortions that had to be reversed if countries were to go back to the market economies of the prewar era. Prices and wages had to adjust, especially if a country were to go back to the Gold Standard, industries that had prospered during war-time autarky would have adjust to international competition, financial institutions would no longer have government guarantees, and would have to learn how to evaluate credit again.

Thus the groups that had benefited most from the recent past had to adjust the most – a situation typical of economies in transition. Nevertheless, some countries went back to the market quite easily. In particular, the United States did not have to depart too much from the market during the war, and also benefited from the enormous purchases made by the Allies, as well as from servicing the export markets the warring parties had abandoned.

Others such as Britain emerged with deficits, debts, and inflation. A mistaken attempt by Britain to go back to the Gold Standard at prewar parities led to further downward pressure on wages and led to the General Strike of 1926. Germany was in an even worse situation because it was saddled with war reparations that could simply not be met. Despite the difficulties of adjustment, however, the prewar liberal consensus was strong enough, as well as the prospects of international trade attractive enough for virtually the entire industrialized world to be on the Gold Standard again by 1927.[67]

But as suggested above, the war had changed the power structure in country after country. Strong interest groups ranging from labor to import-substituting industries were unwilling to bear the risks the market imposed and demanded protection. Government subsidies increased, as did deficits. These deficits could be sustained under the Gold Standard without painful domestic deflation only if some country were willing to finance them by lending. The United States performed this role only till 1928.

As Eichengreen (1996) suggests, perhaps in order to discourage speculation in the stock market, the Federal Reserve raised interest rates in the United States in the first half of 1928. Lending to foreign countries by the United States plummeted to zero in the second half of 1928, Eichengreen (1996, p. 71). In order to stay on the Gold Standard, countries had to either eliminate deficits or raise interest rates and contract credit, both politically painful policies. The stock market crash of 1929 in the United States and the ensuing Great Depression was the

[67] Eichengreen (1996, p. 48).

proverbial straw. With the single biggest consumer in the world contracting, exports plunged all around the world, further aggravating balance of payment problems in Europe, Japan, and Latin America. Domestic price deflation was necessary to stay on the Gold Standard, and to generate the surpluses to repay debt, but one country's deflation was simply a contraction of the export markets for another country.

Moreover, as prices fell, with domestic debt largely nominal, defaults increased. Losses at financial institutions increased, and the threat of a financial crisis put pressure on central banks to intervene and bail out the system. Again, under the Gold Standard, central banks could simply not lend freely to bail out the banking system without jeopardizing the exchange rate.

With domestic demands for insurance and subsidies increasing, there was increasing pressure for governments to "do something". One of the most attractive features of the Gold Standard was that it allowed governments to finance their deficits by borrowing abroad. In fact, for significant periods in the decades preceding 1913, current account deficits exceeded 10 percent of GDP in Australia, Canada, and Argentina, while in the surplus countries of Britain, France, Germany and Netherlands, net capital outflows touched 9 percent.[68] With international lending virtually negligible, even this incentive to adhere to the Gold Standard disappeared.

Thus with export markets moribund, and no possibilities of borrowing, the constituencies in favor of adhering to the Gold Standard were weakened, and country after country abandoned it. A natural consequence was the external discipline on subsidies and distortions was removed and a number of measures mitigating the effects of the markets – erecting tariff barriers, providing unemployment subsidies, providing subsidized credit, bailing out unviable institutions, increasing government investment – were put in place. The command and control structure only recently disbanded after the war provided an attractive template. In general, the return to the centralized "political management that had been first explored during World War 1 became unmistakable by the mid 1930s".[69]

This top down grab for power was extremely dangerous in the case of some countries that harbored unsettled grievances from the past. In particular, Germany, Italy and Japan started rearming. These powers had even less interest in maintaining the existence economic order, Germany in particular because of its large reparation dues. As Polanyi (1944, p. 244-245) writes

[68] Bordo, Eichengreenand Irwin (1999, p. 28).
[69] McNeill, p. 346.

> ”Germany was now eager to hasten the downfall of traditional world economy, which still provided international order with a foothold, and she anticipated the collapse of that economy, so as to have the start of her opponents. She deliberately cut loose from the international system of capital, commodity and currency so as to lessen the hold of the outer world upon her when she would deem it convenient to repudiate her political obligations. She fostered economic autarchy to ensure the freedom required for her far reaching plans. She squandered her gold reserves, destroyed her foreign credit by gratuitous repudiation of her obligations and even, for a time wiped out her favorable trade balance...”

The command economies of these rogue countries could therefore direct resources effectively towards military build-up. Their neighbors could not respond effectively without themselves adopting more centralized structures.

Markets and centralized power are, in general, incompatible. Markets dance to their own tune, and therefore are not responsive to the dictates of centralized power, especially if unaccompanied by monetary reward. By contrast, centralized power wants to mould the economy in its own image as a hierarchy, so that prices and competition do not interfere with the furtherance of its own agenda. The popular groundswell for insurance, the escape from international discipline, and now the need to focus resources on rearmament further strengthened the legitimacy of centralized power. Instead of merely mitigating the adverse effects of the market, it now worked to narrow its sphere and expand the sphere of hierarchies.

In Japan, for example, this implied favoring financial institutions over arm’s length markets. In 1929, Japanese firms raised 26 percent of their debt from bond markets, and only 17 percent from banks. After Japan went off the Gold Standard in December 1931 and intervention in markets increased, the banks with the acquiescence of the Ministry of Finance, formed a Bond Committee in 1933 which determined which firms could issue bonds, on what terms, and when. All bonds were required to be collateralized, and banks were to serve as ”trustees” for the collateral in exchange for a substantial fee. Giving banks the responsibility for determining firms' right to access the public bond markets was like giving a fox who resided in a chicken coop the right to determine which chickens could leave. The obvious outcome was that a flourishing bond market was killed off. By 1936, bonds were down to 14 percent while bank debt was up to 24 percent of liabilities. By 1943, 47 percent of liabilities were bank debt while only 6 percent were bonds.[70]

[70] These figures are from Teranishi (1994).

Moreover, the number of banks declined dramatically over this period as banks were (often forcibly) merged to eliminate ”inefficient” competition. In the span of a few years, the banking system had become much less competitive, and competition from the bond markets was eliminated. And as the country moved towards war, the banking system came under the administrative control of the government.

More generally, the market does not occur naturally, but is a man-made institution, in need of rules and protection as Polanyi (1944) argues. Unfortunately, there is no focussed constituency for the defense of markets, especially in bad times. In particular the very competitive nature of the market that makes it hard to effect cross-subsidies also comes in the way of accumulating funds for coordinated defense. Thus a fundamental freerider problem affects the production of these rules and this protection. Every market participant takes for granted the existence of the market, and thus has only a small interest in spending resources to protect it. The paradox therefore is that only coordinated hierarchical interests can defend the market. But these domestic interests may not be interested in the genies of competition and discipline unleashed by the markets unless forced or enticed by the outside world. In the 1930s, the outside world had shut down.

Moreover, once a country walls itself off, the outside world has to be very attractive before it opens up again. Unlike markets, entrenched hierarchies have the power to defend themselves. So the market can be in retreat for a long time. Consider Japan again. Once the banks had power, they were unlikely to give it up easily. Despite their best efforts to break up the bank firm combines established during the period of militarization, the post-war American occupying forces could not prevent them reemerging as the Keiretsu or main bank system. Even as Japanese firms invaded the rest of the world in the 1970s, their bond markets remained miniscule. It was only in the early 1980s, as Japanese firms decided to borrow abroad rather than depend on their antiquated financial system that Japanese banks had to loosen their stranglehold. The powers of the bond committee were curtailed. The markets had their revenge as the banks paid the price for years of being shielded from competition, making poor credit decisions and driving Japan into crisis.

The intensity of the anti-market revolt varied country by country, depending on the severity of the economic shock and the country’s inherent propensity to centralize. In particular, the different legal and institutional environments provided a different degree of protection against the anti-market backlash. Countries with a civil law system became an easy prey for the political movements advocating a command and control system. After all, civil law, since its Roman origin, has been

an instrument of the State in expanding its power.[71] Thus, the new political objectives, be they a stronger military or a more substantial redistribution of resources, could easily permeate the legal infrastructure of these countries. The most extreme example is Romania, where the communist government that took power after WWII found no need to change the Napoleon Code, because it could be easily adapted with minor changes to the new regime. Not only did the anti-market reaction have greater ability to change the rules in Civil law countries, it could also do so faster since the legislator makes the law, and he has the power to change them as he sees fit.

Common Law provided a stronger bulwark against centralization. First, the English common law tradition was shaped by the struggle between the Crown and the aristocracy. As a result, it emphasizes restraining the government and protecting the individual, making more difficult the centralization of the economy. For example, during the First World War, while France and Germany became command and control economies to facilitate military production, England still retained the spirit of voluntarism in war production, see McNeill (1981, p. 342-343). Second, Common Law is based primarily on judicial precedent, rather than upon the act of a legislator. Thus, a new political climate is less able to transform the structure of society. For example, Roosevelt's attempt to cartelize the U.S. economy failed not because of political opposition, but because of the resistance of the U.S. Supreme Court, which struck down several key pieces of the New Deal legislation. Third, even though Common Law is hard to change from the center, it evolves at the periphery, and innovates around legislative or administrative roadblocks set up by the center. In our view the decentralized nature of the Common Law system, and how it responded to the political anti-market forces can explain the continued vibrancy of financial markets in Common Law countries even when markets elsewhere declined.

It is always dangerous to try to reduce the richness and complexity of the historical evolution of many countries to one single factor. Obviously, many other concurrent events played a role in the extent of the success of the anti-market revolt, not least, the magnitude of the shock the countries had to face during this period and the proximity of the countries in Continental Europe to those rearming. Also, we have emphasized Common Law as the primary liberal institution that saved markets in Anglo-American economies, while any of the liberal institutions that permeate countries of English origin could have helped restrain the anti-market reaction.

[71] Finer (1997).

We can, however, offer some support to our thesis that the centralization of the legal system makes it vulnerable to political pressure. Such a test is provided by a recent comparative study of the evolution of U.S. and U.K. bankruptcy procedures.[72] Both these countries are based on Common Law. The U.S. Constitution, however, left the right to regulate bankruptcy to the Federal Government, an exception to a system where all corporate governance rules are decided at a state level. The British system has no such provision. This represents the perfect natural experiment of our claim that legal centralization itself leads to greater political interference: two initially identical systems that evolved in similar environments, but which differ mainly in the degree of centralization of the legislative process. Over the years the U.S. bankruptcy procedures departed significantly from its British counterpart, which remained substantially unchanged. Consistent with our claim, these deviations introduced a prodebtor bias, in response to political pressure in that direction.

Conclusions

Our brief review of the historical evolution of financial markets seems consistent with the predictions of the political theory. The reason why at the beginning of the 19th century England emerged as the financial center had primarily to do with its political situation. A fortuitous combination of a dominant elite who could benefit from financial development and a weak and decentralized government that could not interfere (and actually had an interest in promoting development so as to be able to finance itself) lead to the growth in the financial market. The subsequent retreat of markets was due to a convergence of anti-market forces: a bottom-up demand of insurance arising from the masses that for the first time acquired some political power and e top-down desire for national security that required a coordination of natural resources.

The analysis also suggests that while a Common law system protects a market economy, it is neither a necessary nor a sufficient condition for a successful financial market. Instead, our analysis emphasizes the importance that political consensus (or lack of) plays in the development and defense of markets in general and financial markets in particular.

[72] Franks and Sussman (1999).

9 Policy Implications

The fundamental conclusion of our analysis of the causes of financial under-development can be summarized by paraphrasing James Carville's 1992 campaign slogan: "It is politics, stupid!". This extremely simple message, though, carries several important lessons, which might be useful in addressing current and future policy challenges.

First, unlike the logical conclusion from recent studies relating financial development to legal origin, our analysis offers hope to those countries that have not had the good fortune to be colonized by the British. Instead of abandoning a legal tradition (not an easy task even if it did not destroy the specific capital of lawyers), they only have to deal with the politics. Clearly, reforming the political system is not much easier than reforming the legal system. However, it may be relatively easy empowering constituencies in favor of financial development.

This is extremely important for transition economies. It is not sufficient to design the rules appropriately, it is also necessary to strengthen the political support for the market in the longer term. The importance of this point can be appreciated by looking at the different long term effects of privatizations in Russia and Poland. Both plans were well thought out. If anything, the Russian plan could pride itself on using top U.S. legal scholars to design its corporation law. The Russian reform could be considered the ultimate social experiment in which the "right" set of laws are introduced to promote financial development. But the Polish reforms have proved more resilient.

The Polish plan placed the shares of state enterprises in big mutual funds, that could not directly run companies, but could gather large enough blocks to control them. Individual investors were then given shares in the mutual funds. This created a powerful group in favor of market development: mutual fund managers. This group successfully fought for more disclosure and more transparency. It was crucial that the role of outside monitors was given to mutual fund managers, as opposed to bank managers. Bankers have a vested interest in maintaining the market opaque in order to exploit the return to their investment in information.

Undoubtedly, Russia was more difficult terrain to start with. But the problems were potentially compounded by the political constituencies empowered by the reforms. The shares of state enterprises were given directly to the citizens, while the control was handed over to corporate

managers. These managers (former communist *appartchiks* who are still very influential) found themselves in the position of the rich in our parable. They controlled monopolist firms and had all the interest in opposing any move toward a more transparent and efficient market, which would have implied less independence for them and more opportunity for entry for competitors. Not surprisingly, few would associate the terms transparent and good governance with Russian firms, and the financial sector is moribund.

A second implication is that outside forces are extremely important in supporting markets. Outside forces such as financial and industrial investors tend to be more immune to cooptation or threats from the government. They do not benefit from many of the politically motivated internal transfers and subsidies. Not being part of the cozy domestic compact, they rely on transparency and legal enforcement. As a result, their interests coincide with those of markets, and they are more sensitive to attempts to shut it down, imposing some discipline on unbridled central power. In the 1930s and 1940s the anti-market revolt was able to proceed to its natural conclusion primarily because the Gold Standard fell apart. Similarly, the resurgence of financial markets in the 1980s is in large part due to the liberalization of trade and capital movements.

In this respect, the role played by international organizations is ambivalent. On the one hand, they had, and still have, a tremendous effect in promoting the cause of liberalization in many countries. The European Union, for instance, is more promarket than any of its individual member (with the possible exception of the United Kingdom). The World Trade Organization helps prevent tariff wars, which proved to be so devastating in the 1930s. On the other hand, the centralization of political power in these organizations runs the risk of weakening the salutary role of outside forces. In a hypothetical new world order, ruled by a supernational Government, outside forces would disappear. There would be no foreign countries to emigrate to, no foreign market to access for capital, no foreign competitors to enter the market. A supernational Government will find it easy to intervene, cross-subsidize, and regulate when the internal domestic forces demand it. The threat to markets is obvious.

And the threat of an anti-market revolt is always latent, ready to explode at the first sign of a major crisis. In fact, in the wake of last year East Asian crisis, many pundits and Government officials started to invoke severe anti-market measures. This is inevitable, because, as we said, markets tend to destroy many primitive forms of insurance, for which they provide very little substitute. Thus, the severity of a shock is felt much more strongly in a market economy, creating discontent, which can be easily used at the political level. The pain produced by the shock

is very salutary from an economic point of view. It is an essential part of the working of a market economy, which accentuates the need for action, accelerating the recovery. But it is extremely costly from a political point of view, especially because pain is not a political message that sells well. All too often, before the market revives itself, the Government intervenes to diffuse the cost of the crisis, slowing the road to recovery.

The fear of a political backlash should caution us against taking for granted that financial development is unidirectional, and that the generalized consensus in favor of markets is irreversible because of their obvious efficiency. Our historical account shows that in time of crises a political backlash may occur and it can have very long-term consequences. Thus, even in times of prosperity (in fact, precisely in those times) it is wise to generate some forms of safety net, to minimize the strength of an anti-market revolt in a time of crisis. This is beneficial even if it has some cost in terms of efficiency. An interesting example of this is presented by Roe (1998) in his analysis of the different methods of land privatization in the Americas during the 19th century. Argentina sold state land in big lots. This generated large revenues and lead to the creation of latifundia, which at the time were the efficient method of farming. The United States, instead, donated small lots of land to those farmers who were willing to settle in. This form of organization was economically less efficient, but more stable politically. At the first agricultural crisis, Argentinean landlords fired a large number of workers who converged to Buenos Aires in search of jobs. This large mass of unemployed became easy fuel for populist anti-market movements, with consequences that are well known. By contrast, American farmers held on to their piece of land, strongly supporting property rights and guaranteeing the survival of the market economy.

The need for some safety net is even more important today, given the potential sources of political tensions looming in front of us. One is represented by the distributional consequences of technological change. While extremely beneficial, technological change has enormous distributional effects, which are amplified by markets. For example, in the United States the wage differential for college graduates has been increasing for the last thirty years. So has the return to talent. These differences are likely to increase redistributional conflict in the future.

The other potential source of political tension is the progressive aging of the population (at least in the most developed economies), with the consequent need to pay for unfunded pension plans. This need will increase the demand for redistribution and, thus, the demand for a more centralized economy. As we already mentioned, in a command and control economy the costs of redistribution are lower, both from an economic and a political point of view, because cross subsidies become

less visible and easier to implement. Thus, the aging of the population could enhance the pressure for more intervention.

In the face of these threats, the survival of financial markets, and in general of market economies, crucially depends upon the ability to preempt these sources of political tensions.

10 Conclusions

The quest for the ultimate cause of financial backwardness is not another senseless academic exercise, but a first order question, which carries enormous policy consequences. The answer should shape any attempt to reform a financial system.

Our conclusion that politics is the main villain carries both a hope and a warning. The hope is that the benefits of financial development are within every country's reach. And today many forces are joining to ease the path to financial development. Increased international competition is eliminating the monopolistic rents of local producers not only weakening their opposition to financial development, but also transforming them into advocates for a better financial system. The large number of countries that in the last few years have introduced new laws in favor of investors supports this hope.

The warning is that the dramatic progress that financial markets have made in the last decade is not irreversible. This has been a time of expansion. The market as an institution remains politically fragile and the inevitable contraction will highlight new political challenges. Only if promarket forces actively recognize these will researchers at the end of the 21st century not have to wonder what caused financial markets to retreat over the turn of the 20th century.

References

Bencivenga and Smith Bencivenga, (1991), Valerie and Smith Bruce. ”Financial Intermediation and Endogenous growth”, *Review of Economic Studies*, April, 58, 195-209.

Boone, P., A. Breach, E. Friedman and S. Johnson, (1998), ”Corporate Governance and the Asian Financial Crisis, 1997-98, mimeo, MIT.

Brealey, R. and S. Myers, (1996), *Principles of Corporate Finance*, McGraw Hill, 5th edition.

Clark, R.C., (1986), *Corporate Law, Corporate law*, Boston: Little, Brown, c1986.

David, René and John E.C. Brierley, (1978), Major Legal Systems in the World Today: An Introduction to the Comparative Study of Law, London Stevens; New York: distributed in U.S.A. by Free Press of Glencoe.

Easterbrook, F. and D. Fischel, (1991), *The Economic Analysis of Corporate Law*, Harvard University Press.

Engelbourg, Saul and Leonard Bushkoff, (1996), The Man Who Found the Money: John Stewart Kennedy and the Financing of the Western Railroads, East Lansing: Michigan State University Press.

Kroszner, R., (2000), ”Is The Financial System Politically Independent? Perspectives on the Political Economy of Banking and Financial Regulation”, Supplement No. 25 to the Government Inquiry on the International Competitiveness of the Swedish Financial Sector.

La Porta, Florencio, Lopez de Silanes, Andrei Shleifer, and Robert Vishny, (1997), ”The Legal Determinants of External Finance”, Journal of Finance, 52.

La Porta, R., Lopez-de-Silanes, F., A. Shleifer and R. Vishny, (1998), ”Law and Finance”, *Journal of Political Economy*, 106, 1113-1155.

Myers, S., (1977), The determinants of corporate borrowing, Journal of Financial Economics, 5, 147-175.

Powell, Ellis, (1915), *The Evolution of the Money Market 1385-1915*, The Financial New Press, London.

Prescott, E., (1986), ”Theory Ahead of Measurement”, *Minneapolis Fed Quarterly Review.*

Rajan, R., (1998), ”The Past and Future of Commercial Banking Viewed through an Incomplete Contract Lens”, Journal of Money, Credit and Banking, vol. 30, (3), 524-550.

Rajan, R. and L. Zingales, (1995), ”What Do We Know about Capital Structure? Some Evidence From International Data”, *Journal of Finance* (1995), 50, 1421-1460.

Rajan, R. and L. Zingales, (1998a), ”Financial Dependence and Growth”, *The American Economic Review*, (1998), 88, 559-586.

Rajan, R. and L. Zingales, (1998b), ”The Tyranny of the Inefficient: An Inquiry into the Adverse Consequences of Power Struggles”, forthcoming in the *Journal of Public Economics*.

Rajan, R. and L. Zingales, (1998c), ”Which Capitalism? Lessons From the East Asia Crisis”, *Journal of Applied Corporate Finance.*

Rajan, R. and L. Zingales, (1999), ”The Firm as a Dedicated Hierarchy: A Theory of the Origins and Growth of Firms”, http://gsblgz.uchicago.edu.

Rajan, R. and L. Zingales, (2000), ”Financial Systems, Industrial Structure, and Growth”, Supplement No. 23 to the Government Inquiry on the International Competitiveness of the Swedish Financial Sector.

Roe, M., (1994), *Strong Managers and Weak Owners: The Political Roots of American Corporate Finance*, Princeton University Press, Princeton, N.J.

Seligman J., (1995), *The Transformation of Wall Street*, Northeastern University Press, Boston 1995, p.42.

Shleifer and R. Vishny. (1997), ”A Survey of Corporate Governance", *Journal of Finance*.

Tunc, Andre, (1991), Corporate Law in R. Bauxbaum, G. Hertig, A. Hirsch, and K. Hopt, eds.: *European Business Law: Legal and Economic Analysis of Integration and Hamornization*, de Gruyter, Berlin.

Zingales, L., (1995), ”What Determines the Value of Corporate Votes?”, *Quarterly Journal of Economics* 110: 1047-1073.

Is The Financial System Politically Independent? Perspectives on The Political Economy of Banking and Financial Regulation

by

Randall S. Kroszner*

Supplement number 25 to the Government Inquiry on the International Competitiveness of the Swedish Financial Sector.

Views, suggestions, and opinions expressed in this study represent the personal views of the author and not necessarily those of the Government Inquiry on the International Competitiveness of the Swedish Financial Sector.

September, 1999

* Professor of Economics, University of Chicago, Graduate School of Business, 1101 East 58th Street, Chicago, IL 60637, USA.

Table of Contents

Kort sammanfattning (in Swedish)

Syftet med studier i politisk ekonomi är att analysera hur regleringar tillkommer, vilka krafter som påverkar denna process och vilka krafter som verkar för att konservera respektive förändra existerande regleringar. Denna bilaga illustrerar a) att det finansiella systemet inte är oberoende av det politiska systemet, och b) vad detta ömsesidiga beroende betyder för regleringar och regleringsprocessen på det finansiella området. Detta innebär ett annat och kompletterande synsätt på regleringar i förhållande till en mer normativ analys av ”optimal” reglering.

I bilagan diskuteras fem olika motiv till finansiella regleringar och avregleringar. För det första finns det ett *kollektivt motiv* till regleringar där syftet är att maximera den samhällsekonomiska välfärden. Eftersom många regleringar inte kan förklaras på detta sätt behandlas för det andra olika *privata intressen* som drivkrafter i regleringsprocessen. Även om teorin om lobbying från privata intressegrupper kan förklara många former av regleringar har den svårt att förklara trenden mot avregleringar under det senaste decenniet. *Ideologiska skift* bland lagstiftare och väljare blir en tredje förklaring, även om orsakerna till dessa skift är svårfångade. För det fjärde kan den *institutionella beslutsstrukturen* påverka både olika intressegruppers incitament att organisera sig och deras möjligheter att påverka besluten. Slutligen kan politikernas och byråkraternas *budgetmässiga överväganden* motivera regleringar som genererar offentliga inkomster.

Dessa perspektiv underlättar en förståelse för hur teknologiska, juridiska och ekonomiska innovationer har drivit fram de globala finansiella avregleringarna. De finansiella kriserna har också betytt mycket för att driva fram reformer och förändringar av regleringar.

Bilagan avslutas med fem slutsatser om hur den politiska reformarbetet kan underlättas. För det första kan väl genomförda konsekvensanalyser av olika förslag spela en viktig roll för att upplysa allmänheten och beslutsfattare. För det andra kan konkurrens mellan olika rivaliserande intressegrupper minska risken att få regleringar som gynnar vissa grupper på det allmännas bekostnad. För det tredje, kan en reformering av regleringar underlättas om den offentliga beslutsprocessen stimulerar konkurrens mellan olika intressegrupper. För det fjärde skulle en större genomlysning av det statliga deltagandet och tillsynen möjliggöra en bättre utvärdering av statens roll i det finansiella systemet. För det femte kan ett inflöde av utländska aktörer vara en positiv kraft, eftersom dessa

har färre politiska kontakter och mindre möjligheter att skaffa sig fördelar i regleringssystemet.

Summary

Rather than take regulations as given, the political economy approach attempts to provide a positive analysis of how and why regulations evolve as they do and what forces can lead to their durability as well as their potential for change. The focus of this paper will be to illustrate how the banking and financial system is not independent of politics and what implications this interdependence has for understanding regulations and reform. This perspective provides an alternative lens through which to analyze regulation and which is complementary to the traditional normative analysis undertaken by economists studying ”optimal” regulation.

Five complementary positive political economy approaches are outlined and applied to understand the pattern of banking and financial regulation and deregulation. First, the public interest approach emphasizes the maximization of social welfare as the motivation for regulation. Since many regulations cannot be rationalized on such grounds, a second approach emphasizing private interests is then developed. This theory focuses on the strength and organization of special interests groups that compete to lobby for protections and privileges. While the private interest theory is able to explain many forms of regulation, it has only partial success in explaining the recent trend towards deregulation. Changing ideology of legislators and voters provides a third alternative. Pro- and anti- business ideological commitments may play a role, but it is difficult to understand exactly what drives ideological change. Fourth, the institutional structure of policy-making can affect both the incentives of interest groups to organize and their effectiveness in influencing policy outcomes. Finally, budgetary considerations of politicians and bureaucrats can motivate regulations that can help to generate funding for government operations.

These approaches then help to identify technological, legal, and economic innovations that have been driving the global move toward financial liberalization and regulatory reform. Changes in the government securities markets world-wide during the last two decades provide an illustration of the political economy factors at work. The role of financial crises in motivating regulatory reform is interpreted as involving important distributional effects and providing information on the costs of the existing regulatory policies.

The paper concludes with five lessons from the positive political economy approaches for ways to facilitate the policy reform process. First, careful cost-benefit analyses of regulations can play an important role in educating the public and policy-makers. Academic studies and arguments can be thought of as equivalent to technological shocks that can change the productivity and effectiveness of different interest group expenditures. Second, competition among rival interest groups can help to dissipate their efforts and reduce the likelihood of narrow special interest regulation. Third, structuring regulatory and governmental decision-making in ways to encourage competition rather than capture by one particular group also can help to remove obstacles to reform. Fourth, greater transparency and disclosure in government supervision and involvement would permit greater monitoring of the government's role in the financial system. Finally, foreign entry can be a positive force since the outsiders typically will not have the same extent of political connections and be less able to capture the regulatory process for private benefits.

1 Introduction[**]

The structure and regulation of a country's financial markets and institutions are the focus of much policy attention for a number of economic and political reasons. Banks and other financial institutions encourage and collect savings that finance a country's economic growth. By allocating the savings to enterprises and monitoring the use of the funds, these institutions and markets play an integral part of the corporate governance system that ultimately affects the productivity of resources throughout the economy (see Rajan and Zingales 2000a and 2000b). Banks and other financial institutions also play a key role in transmitting the government's monetary and credit policies to the rest of the economy. Parts of the financial sector are effectively regulated as means to provide subsidized credit or services to targeted groups (including the government itself) and to protect particular groups (from, for example, competition, hostile takeovers, or expropriation).

While the economics of financial regulation have been studied extensively (see, e.g., Herring and Santomero 2000 and Kroszner 1998a), the politics have received less – albeit increasing – attention. Rather than take regulations as given, the political economy approach attempts to provide a positive analysis of how and why regulations evolve as they do and what forces can lead to their durability as well as their potential for change. The focus of this study will be to illustrate how the banking and financial system is not independent of politics and what implications this interdependence has for understanding regulations and reform. This perspective provides an alternative lens through which to analyze regulation and which is complementary to the traditional normative analysis undertaken by economists studying "optimal" regulation.

When the infamous American bank robber Willie Sutton was asked why he robbed banks, he replied "That's where the money is". The same might be said for why there is such involvement of the government with the banking and financial system – that's where the money is. This idea

[**] An earlier version of this study was presented at Symposium on the International Competitiveness of the Swedish Financial Industry, organized by the Bank of Sweden Tercentenary Foundation, Stockholm, Sweden, March 25-26, 1999. I would like to thank Erik Berglöf, Yrjö Koskinen, and Jonas Niemeyer for helpful comments.

manifests itself in a number of ways, related to the positive political economy analysis. In the next section, I briefly outline a number of approaches to understanding the political economy of government involvement in the economy and then try to understand why the banking and financial system appears to be particularly vulnerable to politicization.

I then examine the technological, legal, and economic shocks that the political economy approach would suggest are the primary factors that disturb a regulatory equilibrium in banking and financial markets. A case analysis of the world-wide reforms in government securities markets then illustrates the motivations for and outcomes of financial regulatory change. The next section then examines the role of ”crises” in regulatory change, emphasizing the redistributional aspects of crises that result in rapid shifts in the strength of different interest groups. The final section draws some tentative lessons from the political economy approaches concerning how to make more likely the ”incentive compatibility” of ”incentive compatible” regulation. This will also include a brief discussion of the role that academic advocates can play in the process of regulation and its reform.

2 Alternative Approaches to the Political Economy of Regulation

Both policy-reformers trying to effect change as well as researchers trying to develop positive theories of government policy-making have struggled to understand how government intervention and regulation occurs and how and whether it can be subsequently sustained (Rodrik 1996). Five related approaches have been taken to analyzing these phenomena: Public Interest, Private Interest, Ideology, Institutions, and Leviathan. While these approaches are not mutually exclusive, they emphasize different aspects of the interaction between economics and politics. Each captures an important element in the process. I will briefly discuss each and apply them to understand aspects of banking and financial regulation.

2.1 Public Interest

The traditional approach that economists took to explaining the existence of regulation emphasized that regulations exist to correct market failures and protect poorly informed consumers from harm.[1] From this perspective, regulatory intervention occurs primarily to maximize social welfare, so this approach is often called the ”public interest theory” of regulation. Public interest rationales are given for capital regulation and deposit insurance to provide a sound banking system because stability of the financial system can have spillover effects for general macroeconomic performance (e.g., Kaufman and Kroszner 1997). Statutory protections of shareholders and creditors from ex post appropriation and supervisory agencies in like a Securities and Exchange Commission are rationalized on the grounds of investor and consumer protection.

A key challenge to the public interest view is that many forms of regulation have little or no redeeming social value. Entry restrictions that protect banks or other financial institutions from competition, portfolio restrictions that hinder diversification, and geographic restrictions that have prevented expansion within a country or across national borders are

[1] Joskow and Noll (1981) call this normative analysis as positive theory.

generally difficult to rationalize on public interest grounds. Statutes or regulatory procedures that protect incumbent management from many forms of discipline by shareholders or outsiders provide additional examples. Regulation that does not appear to serve a public interest also is ubiquitous in other sectors (see Stigler, 1988).

Virtually all regulation, regardless of whether it may have a public interest rationale, has significant distributional consequences. The parties affected by the regulation thus have an incentive to try to ensure that the government structures the regulation in such a way as to benefit them. A public interest argument often is used to mask the private interests that the intervention serves. Private interests may try to confuse the public debate by providing false or misleading information to make it difficult to discern what policy would improve social welfare (e.g., Kane 1996 and Dewatripont and Tirole 1999).

2.2 Private Interest

The ”private interest theory” of regulation, also called the economic theory of regulation, characterizes the regulatory process as one of interest group competition in which compact, well-organized groups are able to use the coercive power of the state to capture rents for those groups at the expense of more dispersed groups (e.g., Olson 1965, Stigler 1971, Peltzman 1976 and 1989, and Becker 1983). Changes in the size, strength, and organization of interest groups thus provide the key to understanding policy changes. Regulated groups may be sufficiently powerful that they influence the politicians and the regulatory bureaucracy to serve primarily the interests of those subject to the regulation. In other words, the regulated group ”captures” the regulators, hence this is sometimes called the ”capture theory” of regulation.

The incentives for such regulatory behavior may be direct or indirect. Pressure may be exerted directly on politicians, though campaign contributions or votes. The politicians then pass a new statute or pressure the regulators to act sympathetically towards the interest group. Indirect incentives may come through regulators understanding that cooperative behavior may be rewarded with lucrative employment opportunities in the industry after leaving the government, a practice so common in the past with Japanese Ministry of Finance officials that it is euphemistically called *amakudari* or the ”descent from heaven”.

The effectiveness of the interest groups depends upon a number of factors. First, cohesive groups will find it easier to organize and overcome free-rider problems in lobbying for regulations that may

benefit them. Producers of goods and services tend to be more compact and better organized than consumers, so there is a tendency for regulation on net to benefit producers more than consumers (Stigler 1971). The parallel in corporate finance is that incumbent managements tend to be better organized and more effective than dispersed shareholders, so there is a tendency for regulation to benefit incumbent management against small shareholders (see Hellwig 1998 and below). The ability of a group to organize is often inversely related to its size, but many labor unions and trade organizations have been able to develop effective lobbying bodies through carefully crafted incentives that provide a variety of information and support services in return for membership (see Olson 1965).

Second, groups tend to be more effective not only when the benefits are concentrated among group members but also when the costs of the regulation are relatively diffuse. A compact group of potential losers each of whom would experience high losses associated with the regulation will be likely to form a lobby that will try to counteract the original interest group's pressure. Interest groups most directly affected by the regulation may build a coalition to lobby for or against the regulation.[2]

Third, in addition to the diffusion of the costs across different groups, the level of the costs relative to the benefits obtained by the interest group play an important role (Becker 1983). Deadweight loss is defined as precisely the difference between the winner's benefit minus the loser's cost from the change in output generated by the regulation. Factors affecting the "efficiency" of the regulatory or transfer mechanism thus may have an important impact on political outcomes. As the deadweight loss of grows, for example, the losers are losing more for each krona of the winner's gain. When this gap widens, losers have a greater incentive to fight each krona of the winner's gain and the winners have less incentive to fight for each krona of the loser's loss. In other words, when deadweight losses are high, an interest group faces greater opposition to

[2] In addition, groups with completely unconnected interests may form "support trading" or "log rolling" coalitions. Two groups may agree to support each other even if the members of one group are not affected by the regulations that the other wants. Tariffs are a classic case of "log rolling" in which, say, lumber and glass producers support each other's call for higher protection, thereby providing greater support for higher tariffs than otherwise would be (Irwin and Kroszner 1999).

its protective regulation on the margin and hence is less likely to be successful.[3]

Similarly, politicians in electoral democracies are concerned about finding an optimal support coalition to promote their re-election chances, so they take into account the marginal costs and benefits to different groups. The rents generated by regulation in an electoral democracy thus are likely to be spread among different groups, even though one group may be the primary beneficiary (Peltzman 1976).[4] Regulation that protects financial institutions from competition and subsidized government deposit insurance to banks generates rents for this sector that are then partially shared through directed credit allocation.[5]

The private interest theory thus helps to explain why the banking and financial system is particularly susceptible to political influence. The banking system provides an effective but off-balance-sheet way for the government to redistribute resources (Kroszner forthcoming). Few if any other sectors provide the same degree of flexibility to redistribute resources, whether implicitly through Bank of Japan ”window guidance” or explicitly through statutes such as the Community Reinvestment Act. Credit allocation through financial institutions can be an important implicit or explicit part of a government's industrial policy.[6] Banks and financial institutions may be induced to act, at least in part, as implicit fiscal arms of the state, but must be compensated through protective regulation.

[3] Becker (1983) argues that competition among lobbying groups thus will lead to the most efficient (lowest deadweight cost) regulations being chosen, so there is a tendency for regulation to be ”efficient” in this sense. Wittman (1995) takes this argument further to conclude that both democratic institutions and outcomes are efficient.

[4] When the constraint of future elections is less binding on politicians, they may engage in less rent-sharing and provide windfalls to targeted groups. McGuire and Olson (1996), however, argue that less democratic regimes may be better able to insulate themselves from rent-seeking and might find it in their own interest to pursue economic policies in the public interest.

[5] Also, flat rate deposit insurance tends to subsidize the smaller and riskier banks at the expense of the larger, better diversified, and safer banks. Lobbying for flat rate deposit insurance historically has been consistent with this pattern of relative benefits (e.g., Calomiris and White 1994 and Hubbard et al. 1996).

[6] Gershenkron (1962), for example, argued that the German government fostered the development of strong universal banks in Germany, at the expense of financial market development, to promote rapid economic development in the nineteenth century.

Since the government is so heavily involved in banking, it may be very difficult to have effective government regulation of the domestic banking and financial sector. In these circumstances, simply hiring more and better-trained supervisors and adopting good regulatory principles is not sufficient because the government may have little incentive to enforce rules of sound banking, either on state-owned banks or privately-owned banks. The co-dependence of the banks on the government and of the politicians on the banking industry allows problems to grow unchecked, as the depth of the banking troubles in the Asian currency crisis countries illustrates. This linkage also may help to explain why governments cannot seem to avoid bail-outs of the financial sector, even as officials acknowledge and decry the moral hazard problems of the bail-outs themselves. These perverse incentives are not unique to developing countries, as the long delays in responding to the Savings and Loan crisis in the U.S. and the banking problems in Japan also show.

In developing a political economy explanation of practices and regulations in corporate finance, Hellwig (1998) emphasizes the contrasting interests and organizational costs facing insiders versus outsiders. He explains, for example, the heavy reliance in firms on internally generated cash flows, rather than external sources of funds, to finance investment as arising not from failures in the capital markets but from the incentives of managers to avoid subjecting themselves to outside discipline. From this perspective, investment choices could be distorted by managers taking into account the potential loss of their private benefits associated with using outside sources of finance.

Insiders, such as firm management and controlling shareholders, have incentives to try to insulate themselves from external pressures and to attempt to disenfranchise other shareholders (Hellwig 1998). Incumbent management and controlling shareholders of large and well-established firms, for example, have been effective in most countries in obtaining statutory and regulatory barriers to raise the costs of hostile takeovers. Part of the strong backlash against Michael Milken and ”junk” bonds, for example, may be because Milken and these instruments permitted small and unknown outsiders to subject even the largest firms to the pressures of the market for corporate control (see Fischel 1997). In addition, the ubiquity of laws that make it difficult for small shareholders to remove directors and exercise other forms of control over management’s decisions can be seen as arising from the greater cohesion and lower costs of organizing incumbent management relative to dispersed shareholders (see La Porta et al. 1999).

2.3 Ideology

While the private interest theory has had much success in explaining a wide variety of regulatory interventions that are difficult to rationalize on public interest grounds, it has been less effective in explaining the widespread economic deregulation that has taken place in many countries during the last two decades (Peltzman 1989 and Noll 1989 but see Kroszner and Strahan 1999). Many political scientists and some economists emphasize the importance of beliefs and ”ideology” of voters and politicians to explain regulation and deregulation (e.g., Kalt and Zupan 1984, Goldstein 1988, and Poole and Rosenthal 1997). Difference across countries or among citizens over time in their general beliefs about the appropriate role of the government in economic affairs might affect the extent of intervention. Roe (1994), for example, has argued that populist fears of excessive concentration of power in the hands of financial elites was an important driving force behind many banking and financial regulations in the early part of this century (but see Hellwig 1998 for an alternative interpretation).

Poole and Rosenthal (1997) have developed a useful measure of ideology based on roll-call voting that rates legislators on a simple left-right scale. This ideology measure has had much success in accounting for a wide variety of economic regulation and deregulation not well explained by private interest group variables or party politics. Berglof and Rosenthal (1999), for example, analyze bankruptcy law in the United States and find that this measure of ideology is a key element for understanding the voting patterns on bankruptcy legislation during the last two centuries. Poole and Rosenthal (1993) find an important role for ideology in the battles over the origins of economic legislator in the United States during the nineteenth century.

Identifying the driving forces behind changes in ideology over time, however, have been difficult. What constitutes ”ideology” and whether it can be measured independent of economic interests is the subject of an extensive and ongoing controversy (see Peltzman 1984 and overviews by Bender and Lott 1996 and Poole and Rosenthal 1996).

2.4 Institutions

The new institutional economics approach emphasizes transactions costs and institutional arrangements for decision-making as key factors influencing the outcome of the policy process (e.g., McCubbins, Noll, and Weingast 1988, North 1990, Williamson 1996, Alston, Eggerston, and North 1996, Dixit 1996, and Irwin and Kroszner 1999). This

approach examines how alternative policy-making structures (e.g., delegation to an independent agency versus a parliamentary vote versus an executive order) influence the incentives of both special interests and governmental actors to shape policy. Opportunities for vote-trading and issue-linkages, for example, may differ under alternative structures and can confer advantages (e.g., agenda control) to particular players. These institutional and transactions costs features can in turn affect the incentives for interest groups to organize and the effectiveness of their lobbying efforts. Interest group size and strength, thus, is not given but may be endogenous and it is important to take such considerations into account if one wishes to make a durable policy change (e.g., Irwin and Kroszner 1999).

The regulation of bank powers illustrates the endogeneity of interests with respect to the regulatory framework (e.g., Kroszner 1996). In 1933, the U.S. adopted the Glass-Steagall Act which fragmented the U.S. financial system by strictly limiting the powers of commercial banks. In particular, commercial banks could not engage in securities underwriting, much in contrast to classic German universal banks and banks in many parts of Europe. While there does not appear to be an economic justification for such a separation (see Kroszner and Rajan 1994 and 1997), there may be a redeeming feature in terms of the political economy of financial regulation.

The silver lining in the cloud of Glass-Steagall is that in the U.S. a rich variety of alternative financial services providers have developed and they compete in both the financial market and the market for financial regulation. In Germany, for example, the early implicit state fostering of strong, universal banks allowed them to capture the regulatory system and thwart the development of alternative institutions and markets. In the U.S., well-organized groups have helped to establish competing regulatory bodies that are likely to keep the market for financial regulation far from being a monopoly even if the Glass-Steagall Act is relaxed today. The financial services sector is the largest source of political action committee (PAC) campaign contributions in the U.S., giving roughly 20 percent of total PAC contributions, but most of these funds are spent on battles among the rival interests rather than on battling the consumer (Kroszner and Stratmann 1998). The initial regulation and the existence of the regulatory bodies helped to provide incentives for the alternative groups to organize and lobby.

2.5 Leviathan

Politicians and the bureaucracy may be considered a distinct interest group concerned about expanding their size and influence over the economy. Niskanen (1971) and Brennan and Buchanan (1977) suggest that an objective of the government may be to maximize or, on the margin, increase its size and expenditures and discuss institutional structures that can mitigate the tendency toward growth. This view has been characterized as the ”Leviathan” approach.

The fiscal demands of the government help to explain some of the close relation between politics and the banking and financial sector and the origins of numerous regulations. Geographic restrictions on banks in the U.S., for example, arose in the early nineteenth century as a way for state governments to maximize revenues from the sale of bank charters by providing a series of local monopolies (Kroszner 1997). The federal government began to grant bank charters during the U.S. Civil War to create a new class of banks that would hold federal debt and, thereby, facilitate the financing of the war effort. The Bank of England was founded as a way to aid in the financing of the Crown in England. More recently, as governments have come to rely more heavily on deficit financing through the issue of sovereign debt, reforms of the government securities markets around the world can be understood from this perspective (as will be discussed below).[7] Debt moratoria, debt abrogation, and changes bankruptcy law also can be seen in this light (e.g., Berglof and Rosenthal 1999, Bolton and Rosenthal 1999, and Kroszner 1999).

[7] The government also raises revenues thorough seniorage, and the ability to tax through inflation is another reason for the government’s long involvement with money and banking affairs.

3 Political-Economy Factors Driving the Recent Trend toward International Financial Liberalization and Regulatory Reform

The interest-group competition framework outlined above suggests a number of factors to explore when trying to understand how financial reform can become politically feasible. In this framework, technological, legal, and/or economic shocks must occur in order to change the relative strengths of interest groups that would then alter the previous political-economy equilibrium. Identifying and analyzing these factors can provide a basis for facilitating and perhaps shaping future reforms (see Kane 1984, 1987, and 1988, Kroszner 1997a, and Kroszner and Strahan 1999).

Technological change is often cited as a key force behind the innovations in financial markets and institutions during the last two decades. In the political-economy framework, technological improvement does more than simply shift the production possibility frontier for an industry. Technical change can have significant distributional consequences, completely independent of its effects on the costs and efficiency of production, that is, such change is rarely ”distributionally neutral”. New products and markets bring forth new constituencies. Innovations affect the pre-existing markets and institutions and cause shifts in the interests and alliances. Changing the relative strength of competing interests can then lead to regulatory reform.

From the political-economy perspective, we must try to identify shocks to the old equilibrium that would lead to regulatory reform fostering globalization and liberalization (see Kroszner and Strahan 1999). A number of shocks, for example, have increased the elasticity of the supply of investors’ and depositors’ funds and thereby have eroded the value of regulation protecting geographic monopolies, whether local or national. First, the invention of the automatic teller machine (ATM) in the early 1970s was one factor that began to reduce the value to the local banks of geographic protections. In countries like the United States, legal challenges about whether an ATM constituted a branch slowed the spread of the ATMs until the courts determined that an ATM was not a

branch, thereby permitted the growth of interstate ATM networks. ATM networks then rapidly spread worldwide.

Second, consumer-oriented money market mutual funds and accounts offered by investment banks arose in the last two decades. These types of new opportunities for individuals demonstrated that banking by mail and telephone, using toll free numbers, provided a feasible and convenient alternative to local banks. Third, technological innovation and deregulation have reduced transportation and communication costs, particularly since the 1970s, thereby lowering the cost for customers to use distant and foreign banks.

Since the increasing elasticity of deposits supplied to banks reduces the value of geographical restrictions to their traditional beneficiaries, these beneficiaries had less incentive to fight strenuously to maintain them. Also, as elasticities increase, there are fewer rents to share among competing groups so regulation becomes less likely (Peltzman 1989). While any deregulation that eliminates inefficient regulation is broadly consistent with the public interest theory, the timing of the deregulation is difficult to explain by that approach. The opening of banking markets occurs precisely when the geographic restrictions are becoming less burdensome for the public, due to the elasticity-increasing innovations discussed above.

On the lending side, increasing sophistication of credit-scoring techniques, following innovations in information processing technology, financial theory, and the development of large credit data bases, has begun to change the relationship-character of bank lending towards less personal and more standardized evaluation. As a result of these innovations, for example, securitization of mortgages, loans, and consumer credits have become commonplace in the developed countries and are becoming increasingly so in emerging markets. In recent years even banks' lending to small businesses has become increasingly automated, relying less on the judgement of loan officers and more on standardized credit scoring programs.

Technological change thus has diminished the value of specialized local knowledge that long-established local bankers might have about the risks of borrowers in the community. Such changes have increased the feasibility and potential profitability for large and foreign banks to enter what had traditionally been the core of small, local bank activities. The large and foreign banks have therefore had an incentive to increase their lobbying pressure to attain the freedom to expand into these markets. In terms of the Becker (1983) model, the deadweight costs of preventing the large and foreign entry is increasing, so the small, local banks are less likely to be able to maintain the restrictions. In addition, as the value of a local banking relationship declined, local firms that were the main

borrowers from the local banks also would be more likely to favor the entry of large and foreign banks into local markets (Kroszner and Strahan 1999).

The method of opening up of the banking markets also is consistent with the private interest theory (see Kroszner 1997). Typically, new foreign entry is first permitted through investment in existing banks and mergers, rather than de novo entry (particularly of institutions that are in financial distress, as Citicorp's new entry into Mexico illustrates). By removing the geographic barriers in this way, the small, local banks have an opportunity to share in the benefits of deregulation by selling out at a premium rather than being competed out of existence. The smaller banks in the country thus would tend to lobby for foreign entry through mergers because they would prefer to have more potential bidders in the market, which tends to increase the premium paid for small banks (Brickley and James 1987).

An increase in foreign financial institution penetration can generate a virtuous circle in that foreign banks tend to be less politically connected and less likely to be able to "capture" the regulatory authorities. In addition, they are less likely to succumb to pressure for directed lending by the government. With capture less likely and fewer direct benefits to the politicians of bank regulation (e.g., through quid pro quos for directed lending), regulatory reform becomes more likely. New Zealand, for example, began its reform process when roughly 30 percent of the banking system was already foreign owned. By the end of the reform process, a very large fraction of the banks had become foreign owned. This helps to increase the likelihood that the reforms are sustainable and not simply temporary. In sum, technological change was a shock to the old political-economy equilibrium and had important distributional consequences that are typically ignored in economists' emphasis on efficiency issues but are extremely important to a positive explanation of regulatory change.

4 Public Finance Motives behind Financial Reform and Globalization: The Case of Government Securities Markets Reform

Since the late 1970s, there has been dramatic growth in the use of publicly-traded debt as a financing tool for both emerging as well as developed countries (Kroszner 1997b). Before this time most countries, with the exception of the US, typically placed a large share of their debt with domestic banks, either directly or through a bank syndicate arrangement.[8] While the banks were to some extent captive financers of the government, they typically received compensation through protective regulation, below market discount loans from the central bank, and implicit lender-of-last-resort or deposit insurance subsidies.[9] With government debt growing much more rapidly than bank assets, however, it was no longer feasible for governments to rely so heavily upon direct funding by the banks.

Motivated by a desire to keep financing costs on their rapidly mounting debt relatively low, politicians thus had incentives to broaden their sources of funding. Consistent with our deadweight cost analysis above, politicians would like to engage in their redistributive activities but must take into account the losses associated with the transfers. The worldwide reforms of the structure and operation of government securities markets during the last two decades, particularly in emerging markets, can be explained in terms of this motive. Auctions replaced or significantly supplemented the traditional placement of securities with

[8] This form of financing could be seen as a mild form of financial repression (Fry 1997).

[9] There is a long and rich history linking a government's financing desires and financial regulation. During the first fiscal revolution in the U.S. when state governments began to rely heavily on debt financing in the 1840s, for example, many states adopted ”free banking” statutes. This legal change eased entry into banking but required the banks to hold state government securities as reserves, thereby boosting the demand for the state's bonds (see Kroszner 1997a and Kroszner and Strahan 1999).

the banks. Simultaneously, the government created or formalized a primary dealer system in which it authorizes specially designated dealers to have the exclusive right to bid directly in the auctions and to have the responsibility of distributing the securities to investors.

An important feature of these reforms was that foreign-controlled financial firms were permitted to enter the market and become primary dealers, thereby encouraging the globalization of the investor base.[10] Previously, developing countries typically had shielded their domestic banking and financial markets from foreign competition. The politicians had a strong incentive to broaden their investor base to finance their growing deficits, and the percent of government debt owned by foreigners has grown rapidly during the last two decades (see Kroszner 1997b).

The internationalization of the government debt markets also has been associated with an increase in the liquidity of these markets. The primary dealers have an obligation to the government, as well as their own private incentive, to foster the growth of a liquid secondary market in government bonds. Liquid secondary markets help to reduce the government's financing costs, by fostering demand by investors (especially foreign investors) who are more willing to hold instruments which have easily observable market prices and can be easily traded. Liquidity also facilitates the dealers' distribution of the securities to investors. The depth of the government securities markets typically has been associated with an increase in the depth and development of other securities in these countries and increased foreign involvement.

As the economic theory of regulatory reform would suggest, the changes which began the opening of domestic financial markets to foreign competition and created liquid debt markets, providing another form of competition to the banking sector, did not occur without some quid pro quo for the banking industry. In particular, the choice of auction technique illustrates a role for private interests in the details of the institutional changes.

Governments consistently adopted sealed-bid, multiple-price auctions (also called ”discriminatory” or ”first-price” auctions) rather than uniform-price auctions (also called ”non-discriminatory” or ”second-price” auctions).[11] The popularity of the multiple-price auction technique

[10] See Drazen (1997) for a detailed political-economy explanation of why governments may wish to sell their debt to foreigners.

[11] In multiple-price auctions, winning bidders pay the price that they bid, so different winners may pay different prices. Winners are determined by ordering bids by price and filling bids from highest to lowest price until the total quantity of securities auctioned has been sold. In a uniform-price

in the recent reforms contrasts sharply with the sustained academic criticism that this format has received relative to the uniform-price technique as a way to issue securities (e.g., Friedman 1960 and Smith 1966, and more recently Rheinhart 1992, U.S. Treasury et al. 1992, Umlauf 1993, Tenorio 1993, and Nyborg and Sundaresan 1996). Unlike the uniform-price auction, the multiple-price auction is subject to the winner's curse. The probability of winning is positively related to the price that one bids in both types of auctions. In the multiple-price auction, unlike in the uniform-price auction, the expected profit from winning is negatively related to the bid price, given that there is some uncertainty as to the exact value of the securities in the secondary market. As a result, bidders will tend to bid a bit less at the auction than what they estimate the secondary market value will be.

In addition, potential bidders without access to detailed information on which to base the estimates of the secondary market value would be less willing to participate directly in a multiple-price than a uniform-price auction. Consequently, the demand curve at auction using a multiple-price format will be below that in the uniform-price auction. The demand curve also is likely to be flatter, since the uncertainties generated by the ”pay what you bid” format tend to make the bidders at the auction more price sensitive.

In principle, the revenue loss from the downward shift in demand at a multiple-price auction relative to a uniform-price auction could be offset by the ability to price-discriminate in the multiple-price auction. Actual and experimental evidence, however, generally indicates that the added revenue from price-discrimination is not sufficient to compensate for the lower and flatter demand curve (see, e.g., Smith 1966). In Mexico, for example, Umlauf (1993) showed that the government’s auction revenue increased in their Treasury bill market when Mexico temporarily switched from multiple-price to uniform-price format. Tenorio (1993) found similar results for Zambia.

The sealed-bid, multiple-price technique also suffers from the potential for manipulation and may foster cartel-like behavior among dealers. The potential for precisely such manipulations was widely understood, having been described by Friedman (1960) decades earlier. When Mexico briefly switched from a multiple-price to a uniform-price auction, for example, bidders’ overall profits fell sharply and auction

---

auction, all of the successful bidders pay the same price. Who wins is determined the same way as in the multiple-price auction, but the price that the winners pay is highest unsuccessful bidder's price, not the price each winner bid (see U.S. Treasury et al. 1992).

revenue rose, suggesting that the multiple-price format permitted greater scope for manipulation (Umlauf 1993).

Given that the potential problems of the multiple-price auction were well-known, why have the reforms almost universally adopted this format? One explanation is that other countries were simply copying the U.S. which had used this format for many years. This solution, however, is unsatisfactory. Although the reforms followed the general pattern of moving in the direction of a US-style market, there are enough country-specific variation that adopting a different auction technique certainly would have been feasible.

An alternative explanation is that the multiple-price technique enhances the value of the information to which the primary dealers have privileged access. The reforms initially gave primary dealers or syndicate members exclusive access to the inter-dealer brokers, consultations with the Ministry of Finance, regular dealings with the central bank through open market operations. Most trading in the government securities also is concentrated in their hands. In a uniform-price auction, information gathered from such sources and activities is less valuable since both the informed and the uninformed bidders will pay the same ”consensus” price. The primary dealers, *ceteris paribus*, thus would prefer to have the government use the multiple-price technique, and governments appear to have obliged. Also, it is extremely difficult to measure the extent of this benefit for there is no line item in the government's budget to represent it. Obscure transfers are much more likely to avoid public scrutiny (as described in more detail at the end of the next section) and, hence, are a preferred means of compensation by the government. The reforms provided some benefits to the government, reducing their fiscal burden, and preserved some rents for the large financial institutions.[12] Part of the trend toward globalization and financial liberalization, thus, can be accounted for by public finance motives.

[12] Rapid technological innovation, however, has begun to erode the information advantages associated with being a primary dealer. Proliferation of inter-dealer broker screens and the growth of organized derivatives markets, for example, are narrowing the information gap between the primary dealers and others. As this trend continues, the value to the primary dealers of the multiple-price format may fall sufficiently that they would be indifferent between the two techniques. Eventually, governments then may switch over to uniform-price auctions and relax some of the distinctions between dealers and non-dealers that no longer provide important benefits to the primary dealers (see Kroszner 1997b).

5 What is the Role of Crises in the Political-Economy of Regulatory Reform?

Reforms are often associated with banking and economic crises, and the ”crisis” hypothesis provides an alternative to the political-economy approach. Developing as well as developed countries experiencing major bank insolvencies have subsequently undertaken some reform and restructuring of their banking regulatory and supervisory systems (Caprio and Klingebiel 1996), and Sweden is no exception. First on the list of sixteen hypotheses about reform drawn up by John Williamson (1994), distilled from the experiences of top policy-makers presented at a conference on ”The Political Economy of Policy Reform,” is that ”policy reforms emerge in response to crisis”

Are crises an independent factor which can be said to ”cause” reform to occur? Rodrik (1996) has been critical of the crisis hypothesis because it is almost nonfalsifiable – if reform does not occur, proponents of this view will say that the crisis was not sufficiently severe – and because reforms responding to similar crises take very different forms (e.g., Caprio and Klingebiel 1996). The U.S., for example, responded to the banking and economic crisis of the early 1930s by fragmenting the financial system (Kroszner 1996). The Glass-Steagall Act of 1933 narrowed the range of activities permissible for commercial banks, and a series of Acts starting with the Federal Home Loan Bank Act of 1932 created modern Savings & Loan institutions which narrowly focused on the financing of residential mortgages. In continental Europe, however, a number of countries responded in the oppose way by increasing the diversification of their financial institutions, by introducing or broadening powers.

From a political-economy perspective, crises are associated with reform because crises are likely to upset the old political-economy equilibrium. There are four reasons for this. First, crises rarely affect all parties similarly and tend to have important distributional consequences. Since the relative position of competing interests is one of the key elements to a political-economy equilibrium, it is thus not surprising that reforms often occur following crises. Powerful groups or coalitions may fragment as their interests diverge during economic trouble, and new constituencies may be created. Although smaller, less diversified banks

tended to support federal deposit insurance, for example, they became politically powerful enough to enact it only in 1933 (Hubbard et al. 1996, but also see Calomiris and White 1994).

Second, economic upheaval can change the relative costs and benefits of particular regulations. An interest rate ceiling, which may act like a price-fixing arrangement among banks to enhance their profits during normal times, for example, could lead to large outflows of funds and liquidity problems during high-interest crisis periods (see Barth 1991 and Kroszner and Strahan 1996). Hyperinflation crises turn many of the regulations that had protected banks from competition into obstacles in the new circumstances. Innovations in financial technology may create new markets and institutions, and new constituencies with them.

Third, crisis can also affect bureaucratic incentives for regulatory change. Deposit insurance, for example, commits the government to bail-out banks that have liquidity and solvency problems. During times of crisis, deposit insurance funds typically are bankrupt so an explicit taxpayer-financed bail-out would be necessary. To postpone such actions, politicians and regulators may have incentives to reduce various regulatory barriers as a quid pro quo for a financial institution using its private funds to bail out a troubled institution. Special dispensations to cross geographic or product lines have occurred in the U.S., particularly during the Savings & Loan crisis (Kroszner and Strahan 1996), Mexico, where Citicorp recently took a large stake in a troubled local Mexican bank, dramatically easing the expansion of its operations in Mexico, and in Japan, where ”arranged” mergers have helped some banks expand into new activities.

Finally, the enormous costs of a financial crisis may serve an important educational role for the public (see Kane 1996). During normal times, individual voters may not know the full value of the implicit or explicit guarantees that the government, that is, the taxpayer, is making. After a crisis, however, the government is likely to have to raise taxes and sell bonds in order to pay for the bail-out. This more explicit accounting will reveal the costs of policies that the public may not have known were so costly. Bank failures thus may heighten the public’s awareness of the costs of regulation and may make it more difficult, that is, more costly in terms of votes, to maintain the old regulatory regime. The banks now would have to provide more support to politicians, for example, through greater campaign contributions, in order to offset the greater popular opposition. Since the banks are experiencing financial distress, they may not be in a strong position to provide the additional funds, so the likelihood of reform increases.

The reform and repeal of the Argentine deposit insurance system follows this pattern. During the 1980s, Argentina experienced two major

banking crises. The first in 1980-1982 has been estimated to have required more than 50 percent of GDP to resolve and the second crisis in 1989-90 roughly 13 percent of GDP to resolve (Rojas-Suarez and Weisbrod 1996 and Lindgren, Garcia, and Saal 1996). With such large costs to the bail-outs, the public was now acutely aware of the costs of government guarantees of deposits. Due to the hyperinflation, there were relatively few deposits left in the bank system to be insured by 1990, so there were fewer depositors demanding insurance. Also, the banks were in a rather weak position. In other words, the crisis involved a dramatic shift in the relative strength of the groups supporting and opposing deposit insurance. In these circumstances, it became politically feasible to eliminate deposit insurance and Argentina did so.

This reform, however, was not completely sustained. Five years later, during the Tequila crisis, the interests in favor of deposit insurance grew and a private deposit insurance scheme was instituted (Guidotti 1996). The deposit insurance premia are relatively high but the insurance agency is owned by the banks that contribute to it. Thus, if the system stays healthy, they earn the profits from the insurance agency but will bear the burdens when the banks require bail-outs.

The changes in geographic restrictions within the U.S. also can be understood within this framework. Kane (1996) argues that an important shock to the old equilibrium favoring branching restrictions was an increase in the public's understanding of the costliness of having government-insured but (geographically) undiversified financial institutions. During the 1980s, an increasing number of depository institution failures and the Savings and Loan crisis culminating in the taxpayer bail-out heightened the awareness by the public of the costs of restrictions that make depository institutions more fragile and more likely to require infusions of taxpayer funds. The result is the 1994 Riegle-Neal Interstate Banking and Branching Efficiency Act which phases out geographic restrictions on bank expansion within the U.S. (see Kroszner and Strahan 1999).

6 Conclusions for Facilitating the Policy Reform Process

The thrust of the arguments above focus on interest group competition and how the battle among the interests will be a key determinant of the regulatory outcomes. Are careful scholarly analyses of regulations and their reforms then of little relevance for policy so academics should retreat to the Ivory Tower? An organized interest group, money, and/or votes may be necessary for a view to prevail in the political marketplace but it is not sufficient, due to the rivalry among interest groups. Theory and facts, not only money and power, are relevant to the debates. Without an interest group to champion a position, however, an argument may have little effect. (Television, radio, and the internet, however, have been reducing the costs for both information to be disseminated and for groups to organize.)

A logical and empirically supported argument affects the productivity of the lobbying efforts by an interest group, much like a technological shock that can increase the productivity of investments. Although rival interests will always have an incentive to provide disinformation and generate ”hack” alternative studies, a well executed can be a great help to a particular group. In terms of Becker’s pressure group model, careful studies can inform the rival groups about the size of the dead weight losses involved with policy alternatives. Reducing uncertainty about the outcome can energize the losing side to increase lobbying effort. Thus, one implication of the political-economy analysis is that the education of the public and of policy-makers of the actual and potential costs of regulation can play a useful and important role in the policy reform process.[13]

[13] Kane (1996) argues that bank regulators and beneficiaries of restrictions on geographic expansion of banks purposefully misinformed the public and legislators about the costs of the regulations. Only a combination of large failures and costly bail-outs with academic studies explaining why the bail-outs are so costly were able to change the perception of the social welfare effects of the regulation. Jensen (1991) argues that much popular support for corporate governance regulation protecting incumbent management arises primarily from ignorance rather than purposeful misinformation, so that more policy-relevant research is important to effect reform.

A second implication is that competition among rival interest groups can increase the likelihood of beneficial reform. Rival groups have an incentive to battle each other in addition to battling the consumer. If they dissipate their efforts against each other, they are less likely to be able to support narrow special interest regulation. In many emerging markets today, e.g., Russia, a major question concerns whether creating universal banks would allow one particular interest to have too much political power and thwart reform. In addition, the rival groups have an incentive to try to unmask any misinformation that the competing side is generating. This can help to inform both the policy-makers and the public.

Third, the structure of regulatory and government institutions also plays a role. A clear structure of legislative oversight of the regulatory process through, for example, specific committees in the Parliament with responsibility for banking and financial matters may provide a forum which fosters the information generation process (Gilligan and Krehbiel 1989, McCubbins, Noll, and Weingast 1989, Krehbiel 1991, Austin-Smith and Wright 1992 and 1994, Kroszner and Stratmann 1998). Similarly, the incentives for groups to overcome free-rider problems and organize is related to the expected benefit of them doing so. In other words, the organization of interests is endogenously related to the structure of the regulatory process (see Irwin and Kroszner 1999). Opening the regulatory process to include clear channels for new groups that would tend to oppose narrow special interest ”capture” regulation increases the likelihood of regulatory reform by increasing the costs of maintain the regulation to the special interest.

Fourth, greater transparency in government involvement in the financial system is significant. Politicians often use the financial system, either through implicit guidance or explicit through state owned banks, to provide low-cost financing to targeted industries or groups. Directed lending leads to implicit or explicit quid pro quos in order to have the banking sector follow this direction. Problems in Korean banks, for example, stem from encouragement by the government to continue lending to troubled enterprises in return for implicit assurances of a bail-out. Privatization of state owned enterprises, for example, can reduce the benefit to politicians' of directing credit and can generate new constituencies for an efficient banking and financial sector – as long as the firm has been fully privatized and does not have special influence with the government (Kroszner 1996b). Requiring that any such transfers or subsidies be explicitly included in the government's fiscal accounts would clarify such transactions and help to break nexus of implicit agreements and quid pro quos through regulation that support them. Less secrecy and increased disclosure in government supervision

of financial institutions also would permit increased monitoring of the government's role in the financial system.

Finally, foreign entry can generate a virtuous circle because foreign institutions tend to be less politically connected domestically and less likely to be able to capture the regulatory authorities. Foreign institutions also are less likely to succumb to pressure for directed lending by the government. With capture less likely and fewer direct benefits to the politicians of regulation (e.g., through quid pro quos for directed lending), regulatory reform becomes more likely. While there is no simple formula for successful and sustained banking and financial regulatory reform, a positive analysis of the political-economy of rent-seeking does suggest how process and institutions facilitate beneficial reform.

References

Alesina, Alberto and Allan Drazen, (1991), ”Why Are Stabilizations Delayed?” *American Economic Review*, December: 1170-88.

Alston, Lee, Thrainn Eggertsson, and Douglass North, (1996), eds. *Empirical Studies in Institutional Change*, New York: Cambridge University Press.

Austin-Smith, David and John Wright, (1992), ”Competitive Lobbying for a Legislator’s Vote”, *Social Choice and Welfare* 9: 229-57.

Austin-Smith, David and John Wright, (1994), ”Counteractive Lobbying”, *American Journal of Political Science* 38, February: 25-44.

Bartolini, Leonardo and Carlo Cottarelli, (1997), ”Treasury Bill Auctions: Issues and Uses”, pp. 267-336 in Mario Blejer and Teresa Ter-Minassian, eds., *Macroeconomic Dimensions of Public Finance*, New York: Routledge.

Becker, Gary S., (1983), ”A Theory of Competition among Pressure Groups for Political Influence”, *Quarterly Journal of Economics*: 371-400.

Bender, Bruce, and John R. Lott, (1996), ”Legislator Voting and Shirking: A Critical Review of the Literature”, *Public Choice* 87: 67-100.

Benston, George J. and George G. Kaufman, (1996), ”The Appropriate Role of Bank Regulation”, *The Economic Journal* 106, May: 688-97.

Berglöf, Erik and Howard Rosenthal, (1999), ”The Political Economy of American Bankruptcy: The Evidence from Roll Call Voting, 1800-1978”, working paper, Princeton University, April.

Bolton, Patrick and Howard Rosenthal, (1999), ”Political Intervention in Debt Contracts: Moratoria and Bailouts”, working paper, Princeton University, March.

Brennan, Geoffrey and James Buchanan, (1977), ”Towards a Tax Constitution for Leviathan”, *Journal of Public Economics* 8: 255-73.

Brickley, James A. and Christopher M. James, (1987), ”The Takeover Market, Corporate Board Composition, and Ownership Structure: The Case of Banking”, *Journal of Law and Economics* 30, April: 161-80.

Calomiris, Charles and Eugene White, (1994), ”The Origins of Federal Deposit Insurance”, pp. 145-88 in Claudia Golden and Gary Libecap, eds., *The Regulated Economy*, Chicago: University of Chicago Press.

Caprio, Gerard, Jr. and Daniela Klingebiel, (1996), ”Dealing with Bank Insolvencies: Cross Country Experience”, Policy Research Department Working Paper, The World Bank, February.

Casella, Alesandra and Barry Eichengreen, (1994), ”Can Foreign Aid Accelerate Stabilization?”, Working Paper.

Dewatripont, Mathias and Jean Tirole, (1999), ”Advocates”, *Journal of Political Economy* 107, February: 1-39.

Dixit, Avinash, (1996), *The Making of Economic Policy: A Transactions-Cost Politics Perspective*, Cambridge, MA: MIT Press.

Drazen, Allan, (1997), ”Towards a Political-Economic Theory of Domestic Debt”, in Guillermo Calvo and Mervyn King, eds., *The Debt Burden and Monetary Policy*, London: Macmillan.

Fernandez, Raquel and Dani Rodrik, (1991), ”Resistance to Reform: Status Quo Bias in the Presence of Individual-Specific Uncertainty”, *American Economic Review*, December: 1146-55.

Fischel, Daniel, (1997), *Payback*, New York.

Friedman, Milton, (1960), *A Program for Monetary Stability*, New York: Fordham University Press.

Fry, Maxwell, (1997), *Emancipating the Banking System and Developing Markets for Government Debt*, London: Routledge.

Gerschenkron, A., (1962), *Economic Backwardness in Historical Perspective*, Cambridge, MA: Harvard University Press.

Gilligan, Thomas and Keith Krehbiel, (1989), ”Asymmetric Information and Legislative Rules with a Heterogeneous Committee”, *American Journal of Political Science* 33: 459-90.

Goldstein, Judith, (1988), ”Ideas, Institutions, and American Trade Policy”, *International Organization* 42: 179-217.

Guidotti, Pablo, (1996), ”Debt, Monetary, and Banking Policy in Emerging Markets: Reflections from the ‘Tequila’ Effect,” University of Maryland, Department of Economics Working Paper No. 28, December.

Hellwig, Martin, (1998), ”On the Economics and Politics of Corporate Finance and Corporate Control,” working paper.

Herring, Richard and Anthony Santomero, (2000), ”What Is Optimal Financial Regulation?”, Supplement No. 26 to the Government Inquiry on the Internationel Competitiveness of the Swedish Financial Sector.

Holderness, Clifford, Randall Kroszner, and Dennis Sheehan, (1999), ”Were the Good Old Days that Good? Changes in Managerial Stock Ownership since the Great Depression”, *Journal of Finance*, April: 435-69.

Hubbard, R. Glenn, Darius Palia, and Nicolas Economides, (1996), ”The Political Economy of Branching Restrictions and Deposit Insurance”, *Journal of Law and Economics*, October.

Irwin, Douglas and Randall Kroszner, (1999), ”Interests, Institutions, and Ideology in Securing Policy Change: The Republican Conversion to Trade Liberalization after Smoot-Hawley”, *Journal of Law and Economic*, October: forthcoming.

Jayaratne, Jith and Philip E. Strahan, (1996), ”The Finance-Growth Nexus: Evidence from Bank Branch Deregulation”, *Quarterly Journal of Economics* 101: 639-70.

Jensen, Michael, (1991), ”Corporate Control and the Politics of Finance”, *Journal of Applied Corporate Finance* 4: 13-33.

Joskow, Paul and Roger Noll, (1981), ”Regulation in Theory and Practice: An Overview”, pp. 1-65 in Gary Fromm, ed., *Studies in Public Regulation*, Cambridge: MIT Press.

Kane, Edward, (1984), ”Regulatory Structure in Futures Markets: Jurisdictional Competition between the SEC, the CFTC, and Other Agencies”, *Journal of Futures Markets* 4: 367-84.

Kane, Edward, (1987), ”Competitive Financial Reregulation: An International Perspective”, pp. 111-45 in *Threats to International Financial Stability*, edited by Richard Portes and Alexander Swoboda, New York: Cambridge University Press.

Kane, Edward J., (1988), ”How Market Forces Influence the Structure of Financial Regulation,” chapter 9 in Kushmeider, ed., *Restructuring Banking and Financial Services*, Washington: American Enterprise Institute.

Kane, Edward J., (1996), ”De Jure Interstate Banking: Why Only Now?”, *Journal of Money, Credit, and Banking* 28; 2, May: 141-61.

Kaufman, George and Randall Kroszner, (1997), ”How Should Financial Institutions and Markets be Structured?”, pp. 97-122 in Liliana Rojas-Suarez, ed., *Safe and Sound Financial Systems: What Works for Latin America?*, Washington: Inter-American Development Bank.

Klein, Daniel, (1999), ed. *What Do Economists Contribute*?, New York: New York University Press.

Krehbiel, Keith, (1991), *Information and Legislative Organization*, Ann Arbor: University of Michigan Press.

Kroszner, Randall S., (1996a), ”The Evolution of Universal Banking and Its Regulation in Twentieth Century America,” pp. 70-99 in Anthony Saunders and Ingo Walter, eds., *Universal Banking: Financial System Design Reconsidered*, New York: Irwin Professional Publishers.

Kroszner, Randall S., (1996b), ”An Analysis of Bank Privatization”, working paper, Graduate School of Business, University of Chicago.

Kroszner, Randall S., (1997a), ”The Political Economy of Banking and Financial Regulation in the U.S.”, pp. 200-13 in George M. von Furstenberg, ed., *The Banking and Financial Structure in the NAFTA Countries and Chile*, Boston: Kluwer Academic Publishers.

Kroszner, Randall S., (1997b), ”Global Government Securities Markets: The Economics and Politics of Market Microstructure Reforms”, pp. 86-110 in Guillermo Calvo and Mervyn King, eds., *The Debt Burden and Monetary Policy*, London: Macmillan.

Kroszner, Randall S., (1998a), ”Rethinking Banking Regulation: A Review of the Historical Evidence”, *Journal of Applied Corporate Finance*, Summer: 48-58.

Kroszner, Randall S., (1998b), ”The Political Economy of Banking and Financial Regulatory Reform in Emerging Markets”, *Research in Financial Services* 10: 33-51.

Kroszner, Randall S., (1999), ”Is It Better to Forgive than to Receive? Repudiation of the Gold Indexation Clause in Long-term Debt Contracts during the Great Depression,” working paper, University of Chicago, Graduate School of Business, April.

Kroszner, Randall S., ”Less is More in the New International Financial Architecture”, in W. C. Hunter and G. G. Kaufman, editors, *The Asia Crisis: Prospects for Reform*, forthcoming.

Kroszner, Randall S. and Raghuram G. Rajan, (1994), ”Is the Glass-Steagall Act Justified? A Study of the U.S. Experience with Universal Banking before 1933”, *American Economic Review* 84, September: 810-32.

Kroszner, Randall S. and Raghuram G. Rajan, (1997), ”Organization Structure and Credibility: Evidence from Commercial Bank Securities Activities before the Glass-Steagall Act”, *Journal of Monetary Economics* 29, August: 475-516.

Kroszner, Randall S. and Philip Strahan, (1996), ”Regulatory Incentives and the Thrift Crisis: Dividends, Mutual-to-Stock Conversions, and Financial Distress”, *Journal of Finance 51*, September: 1285-1320.

Kroszner, Randall S. and Philip Strahan, (1999), ”What Drives Deregulation? Economics and Politics of the Relaxation of Bank Branching Restrictions”, *Quarterly Journal of Economics* 114: forthcoming.

Kroszner, Randall S. and Philip E. Strahan, (1999), ”Bankers on Boards: Monitoring, Conflicts of Interest, and Lender Liability”, working paper, Graduate School of Business, University of Chicago, March.

Kroszner, Randall S. and Thomas Stratmann, (1998), ”Interest Group Competition and the Organization of Congress: Theory and Evidence from Financial Services Political Action Committees”, *American Economic Review* 88, December: 1163-87.

Kroszner, Randall S. and Thomas Stratmann, (1997), ”Notes on Development of the Congressional Committee System in the Twentieth Century”, working Paper, University of Chicago, September.

Kroszner, Randall S. and Thomas Stratmann, (1999), ”Does Political Ambiguity Pay? Corporate Campaign Contributions and the Rewards to Legislator Reputation”, working paper, Graduate School of Business, University of Chicago, January.

Laban, Raul and Federico Sturzengegger, (1994), ”Distributional Conflict, Financial Adaptation, and Delayed Stabilization”, *Economics and Politic*, November: 257-76.

La Porta, Rafael, Florencio Lopez-di-Silanes, Andrei Shleifer, and Robert Vishny, (1998), ”Law and Finance”, *Journal of Political Economy* 106, December: 1113-55.

La Porta, Rafael, Florencio Lopez-di-Silanes, and Andrei Shleifer, (1999), ”Corporate Ownership around the World”, *Journal of Finance*, April: forthcoming.

McCubbins, Matthew, Roger Noll, and Barry Weingast, (1988), ”Structure and Process, Politics and Policy: Administrative Arrangements and the Political Control of Agencies”, *Virginia Law Review* 78: 431-82.

McGuire, Martin and Mancur Olson, (1996), ”The Economics of Autocracy and Majority Rule”, *Journal of Economic Literature*, March: 72-96.

Miller, Geoffrey, (1998), ”Political Structure and Corporate Governance: Some Points of Contrast between the United States and England”, *Columbia Business Law Review*, 51-78.

Mondino, Guillermo, Frederico Sturzenegger, and Mariano Tommasi, (1992), ”Recurrent High Inflation and Stabilization: A Dynamic Game”, working paper.

Niskanen, William, (1971), *Bureaucracy and Representative Government*, Chicago: Aldine-Atherton.

North, Douglass, (1990), ”A Transactions Cost Theory of Politics”, *Journal of Theoretical Politics* 2: 355-67.

Noll, Roger, (1989), ”Comment on Peltzman”, *Brookings Papers: Microeconomics*, 48-58.

Olson, Mancur, (1965), *The Logic of Collective Action*, Cambridge, MA: Harvard University Press.

Olson, Mancur, (1982), *The Rise and Decline of Nations: Economic Growth, Stagflation, and Social Rigidities*, New Haven: Yale University Press.

Peltzman, Sam, (1976), ”Toward a More General Theory of Regulation”, *Journal of Law and Economics* 19; 1, April: 109-148.

Peltzman, Sam, (1984), ”Constituent Interest and Congressional Voting”, *Journal of Law and Economics* 27: 181-210.

Peltzman, Sam, (1989), ”The Economic Theory of Regulation after a Decade of Deregulation”, *Brookings Papers: Microeconomics*: 1-41.

Poole, Keith T. and Howard Rosenthal, (1993), ”The Enduring Nineteenth-Century Battle for Economic Regulation: The Interstate Commerce Act Revisited”, *Journal of Law and Economics* 36, October: 837-60.

Poole, Keith T. and Howard Rosenthal, (1996), ”Are Legislators Ideologues or the Agents of Constituents?”, *European Economic Review* 40: 707-717.

Poole, Keith and Howard Rosenthal, (1997), *Congress: A Political-Economic History of Roll Call Voting,*, Oxford: Oxford University Press.

Rajan, Raghuram and Luigi Zingales, (1997), ”The Tyranny of the Inefficient: An Enquiry into the Adverse Consequences of Power Struggles”, Working Paper, University of Chicago.

Rajan, Raghuram and Luigi Zingales, (2000a), ”Financial Systems, Industrial Structure, and Growth”, Supplement No. 23 to the Government Inquiry on the International Competitiveness of the Swedish Financial Sector.

Rajan, Raghuram and Luigi Zingales, (2000b), ”Politics, Law, and Financial Development”, Supplement No. 24 to the Government Inquiry on the International Competitiveness of the Swedish Financial Sector.

Reinhart, Vincent, (1992), ”An Analysis of Potential Treasury Auction Techniques”, *Federal Reserve Bulletin*, June: 403-13.

Roe, Mark, (1994), *Strong Managers, Weak Owners*, Princeton: Princeton University Press.

Rodrik, Dani, (1996), ”Understanding Economic Policy Reform”, *Journal of Economic Literature*, March: 9-41.

Romano, Roberta, (1997), ”The Political Dynamics of Derivative Securities Regulation”, *Yale Journal on Regulation* 14, Spring: 279-406.

Rose-Ackerman, Susan, (1978), *Corruption: A Study in Political Economy*, New York: Academic Press.

Sheng, Andrew, (1996), *Bank Restructuring: Lessons from the 1980s*, Washington: The World Bank.

Shepsle, Kenneth and Barry Weingast, (1995), eds. *Positive Theories of Congressional Institution*, Ann Arbor: University of Michigan Press.

Stigler, George J., (1971), ”The Theory of Economic Regulation”, *Bell Journal of Economics and Management Science* 2; 1, Spring: 3-21.

Stigler, George J., (1988), *Chicago Studies in the Political Economy*, Chicago: University of Chicago Press.

Umlauf, S., (1993), ”An Empirical Study of the Mexican Treasury Bill Auction”, *Journal of Financial Economics* 33; 3: 313-40.

U.S. Department of Treasury, U.S. Securities and Exchange Commission, and the Board of Governors of the Federal Reserve System, *Joint Report on the Government Securities Market*, Washington: Government Printing Office, January (1992).

Weingast, Barry R. and William J. Marshall, (1988), ”The Industrial Organization of Congress; or, Why Legislatures, like Firms, Are Not Organized as Markets”, *Journal of Political Economy* 96, February: 132-63.

White, Eugene, (1983), *The Regulation and Reform of the American Banking System, 1900-1929*. Princeton: Princeton University Press.

Williamson, John, (1994), ed. *The Political Economy of Policy Reform*, Washington: Institute for International Economics.

Williamson, Oliver, (1996), ”The Politics and Economics of Redistribution and Efficiency”, In *The Mechanisms of Governance*, Oxford: Oxford University Press.

Wittman, Donald N., (1995), *The Myth of Democratic Failure*. Chicago: University of Chicago Press.

What Is Optimal Financial Regulation?

by

Richard J. Herring*

and

Anthony M. Santomero**

Supplement number 26 to the Government Inquiry on the International Competitiveness of the Swedish Financial Sector.

September, 1999

* Jacob Safra Professor of International Banking; Professor of Finance, Department of Finance, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA.

** Richard King Mellon Professor of Finance, Department of Finance, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA.

Table of Contents

Kort sammanfattning (in Swedish)

Det grundläggande målet för reglering av finansiell verksamhet är att förebygga s.k. systemrisker, dvs. risken att en störning fortplantar sig på ett sätt som de enskilda aktörerna inte kan hantera och som kan ge stora samhällsekonomiska skadeverkningar. Denna typ av risk hänför sig i första hand till bankerna som dels har en sammansättning av sin balansräkning som gör dem särskilt exponerade för denna typ av risk, dels av tradition utgjort den dominerande delen av det finansiella systemet.

De ”skyddsnät” som etablerats genom olika former av reglering har varit framgångsrika i så måtto att svåra systemkriser kunnat undvikas. Samtidigt skapar dock regleringar även problem, bl.a. när det gäller aktörernas incitament ifråga om rikstagande och riskhantering. Dessa problem kan visserligen i viss utsträckning hanteras på olika sätt, men dilemmat består.

Den teknologiska och marknadsmässiga utvecklingen, främst i USA men även i andra länder, gör dock både att bankerna får en mindre dominerande roll i det finansiella systemet, och att bankerna får en annorlunda ekonomisk och finansiell struktur som gör dem mindre avvikande från andra företag och mindre exponerade för systemrisk. Ur samhällssynpunkt är detta en positiv utveckling som bör underlättas, exempelvis genom anpassning av regelverk och tillsyn, eftersom detta innebär att den speciella regleringen av banksystemet, med de negativa bieffekter detta har, då successivt kan avvecklas.

Executive Summary

The financial system is regulated to achieve a wide variety of purposes. However, the objective that distinguishes financial regulation from other kinds of regulation is that of safeguarding the economy against systemic risk. Concerns regarding systemic risk focus largely on banks, because they traditionally have been viewed as occupying a special role in the economy. The safety nets that have been rigged to protect banks from systemic risk have succeeded in preventing banking panics. However, this success has come at the cost of distorting incentives for risk taking. Regulators have a variety of options to correct this distortion, but none can be relied upon to produce an optimal solution.

Technological and conceptual advances may be ameliorating the problem, nonetheless. Banks are becoming less special. As is evident from the data reported here, the US is leading the way. The trends are apparent in other industrial countries as well. Once banks have lost their special status, financial safety nets may be dismantled thus ending the distortions they create. Ultimately, regulation for prudential purposes may be completely unnecessary. The optimal regulation for safety and soundness purposes may be no regulation at all.

The challenge facing regulators is to facilitate these advances and hasten the end of the special status of banks.

Here, we offer several simple prescriptions to improve financial regulation. *First*, the authorities should encourage the introduction of technological improvements that are lowering the costs of information and the costs of storing, retrieving and organizing these data. They should be active supporters of competition in the technology and communication sectors.

Second, regulators should resist the temptation to re-regulate or promulgate regulations that will forestall the inevitable financial restructuring that is part of this change process.

Third, since market discipline will increasingly substitute for prudential regulation, it is important to assure that both regulation and the regulatory staff are of a quality that is consistent with global standards. In terms of the former, increasing emphasis must be placed on market values throughout the regulatory process, and it is important to improve disclosure standards as well. In terms of the latter, the quality and expertise of the regulation and examination staff must keep pace with the escalating standards of the global marketplace.

1 Introduction***

The financial system is regulated to achieve a wide variety of purposes. However, the objective that distinguishes financial regulation from other kinds of regulation is that of safeguarding the economy against systemic risk. Concerns regarding systemic risk focus largely on banks, which traditionally have been considered to have a special role in the economy. The safety nets that have been rigged to protect banks from systemic risk have succeeded in preventing banking panics, but at the cost of distorting incentives for risk taking. Regulators have a variety of options to correct this distortion, but none can be relied upon to produce an optimal solution.

Technological and conceptual advances may be ameliorating the problem, nonetheless. Banks are becoming less special. The US is leading the way, but the trends are apparent in other industrial countries as well. The challenge facing regulators is to facilitate these advances and hasten the end of the special status of banks. Once banks have lost their special status, financial safety nets may be dismantled thus ending the distortions they create. Ultimately, regulation for prudential purposes may be completely unnecessary. The optimal regulation for safety and soundness purposes may be no regulation at all.

*** An earlier version of this paper was presented at the symposium on the International Competitiveness of the Swedish Financial Industry, organized by the Swedish Tercentenary Foundation on March 25, 1999.

2 Rationales for financial regulation

A well-functioning financial system makes a critical contribution to economic performance by facilitating transactions, mobilizing savings and allocating capital across time and space. Financial institutions provide payment services and a variety of financial products that enable the corporate sector and households to cope with economic uncertainties by hedging, pooling, sharing and pricing risks. A stable, efficient financial sector reduces the cost and risk of investment and of producing and trading goods and services.[1]

Financial markets also provide a crucial source of information that helps coordinate decentralized decisions throughout the economy. Rates of return in financial markets guide households in allocating income between consumption and savings, and in allocating their stock of wealth. Firms rely on financial market prices to inform their choices among investment projects and to determine how such projects should be financed.[2]

In view of these critical contributions to economic performance it is not surprising that the health of the financial sector is a matter of public policy concern and that nearly all national governments have chosen to regulate the financial sector. Merton (1990) is undoubtedly correct when he argues that the overall objective of regulation of the financial sector should be to ensure that the system functions efficiently in helping to deploy, transfer and allocate resources across time and space under conditions of uncertainty.

However, actual financial regulation attempts to accomplish several objectives beyond facilitating the efficient allocation of resources. In fact, at least four broad rationales for financial regulation may be identified: safeguarding the financial system against systemic risk, protecting consumers from opportunistic behavior, enhancing the efficiency of the financial system, and achieving a broad range of social objectives from increasing home ownership to combating organized crime.

[1] See Herring and Santomero (1991) for a detailed discussion of the role of the financial sector in a developed economy. For a more recent reference, see Allen and Santomero (1997).

[2] This is the role emphasized by Merton (1989). See LIewellyn (1999) for discussion of the economic rationale for regulation.

2.1 Guarding against systemic risk

Safeguarding financial markets and institutions from shocks that might pose a systemic risk is the prime objective of financial regulation. Systemic risk may be defined as the risk of a sudden, unanticipated event that would damage the financial system to such an extent that economic activity in the wider economy would suffer. Such shocks may originate inside or outside the financial sector and may include the sudden failure of a major participant in the financial system, a technological breakdown at a critical stage of settlements or payments systems, or a political shock such as an invasion or the imposition of exchange controls in an important financial center. Such events can disrupt the normal functioning of financial markets and institutions by destroying the mutual trust that lubricates most financial transactions.

Figure 1. Regulatory measures and regulatory objectives

Regulatory Measures	Systemic Risk	Consumer Protection	Efficiency Enhancement	Broader social objectives
Antitrust enforcement/ ompetition policy		✓	✓	✓
Asset restrictions	✓			✓
Capital adequacy standards	✓	✓		
Conduct of business rules		✓	✓	✓
Conflict of interest rules		✓	✓	
Customer suitability requirements		✓		
Deposit insurance	✓	✓		
Disclosure standards	✓	✓	✓	
Fit and proper entry tests	✓	✓	✓	
Interest rate ceilings on deposits	✓			✓
Interest rate ceilings on loans		✓		✓
Investment requirements				✓
Liquidity requirements	✓	✓		
Reporting requirements for large transactions				✓
Reserve requirements	✓	✓		
Restrictions on geographic reach				✓
Restrictions on services and product lines	✓			✓

Adapted from Herring and Litan (1995).

As an examination of the Systemic Risk column of Figure 1 indicates, a substantial number of regulatory measures have been justified on grounds that they help safeguard the financial system from systemic risk. However, research has shown that a number of these measures, such as restrictions on product lines, are ineffectual at best in safeguarding

against systemic risk and may weaken regulated institutions by preventing them from meeting the changing needs of their customers. Some measures, such as interest rate ceilings on deposits that were intended to prevent ”excessive competition”, may actually exacerbate vulnerability to systemic risk. For example, when interest rate ceilings are binding, depositors will have an incentive to shift from bank deposits to assets yielding a market rate of return thus inducing funding problems for banks.

It should be noted also that some regulatory measures work at cross-purposes. For example, geographic restrictions on banking, intended to protect the access to credit of local firms and households, may increase exposure to systemic risk by impeding diversification of regulated institutions and increasing their vulnerability to a local shock. Similarly, the ”fit and proper tests” one might want to impose for safety and soundness reasons may pose entry barriers that are too high to achieve the efficiency gains from competition. We will examine systemic risk and measures to counter systemic risk in greater detail in sections 3 and 4.

2.2 Protecting consumers

The second fundamental rationale for financial regulation is the protection of consumers against excessive prices or opportunistic behavior by providers of financial services or participants in financial markets. (See the Consumer Protection column of Figure 1.) Antitrust enforcement is the most obvious policy tool to counter excessive prices.

Competition policy is motivated not only by the concern to protect consumers from monopolistic pricing, but also by the aim of harnessing market forces to enhance the efficiency of the allocation within the financial sector and between the financial sector and the rest of the economy.[3]

The United States was the first nation to adopt antitrust policy, which, of course, is concerned with monopolistic pricing in all markets not just financial markets. Over the past decade the European Commission has increasingly taken a more activist role in promoting competition. Last year significant attention was focused on substantial price variations within various categories of financial products offered within the European Union.[4] Although substantial gains have yet to be

[3] See section 2.3 for a further discussion of this point.
[4] See European Commission, 1998.

realized, the European Union's goal of forming a single market in financial services is aimed at increasing competition and lowering prices to users of financial services.

Consumers of financial services – particularly unsophisticated consumers – find it very difficult to evaluate the quality of financial information and services provided to them. In part this is because payment for many financial transactions must often be made in the current period in exchange for benefits that are promised far in the future. But even after the decision is made and financial results are realized, it is difficult to determine whether an unfavorable outcome was the result of bad luck, even though good advice was competently and honestly rendered, or the result of incompetence or dishonesty.

Customers face a problem of asymmetric information in evaluating financial services. Consequently they are vulnerable to (adverse selection), the possibility that a customer will choose an incompetent or dishonest firm for investment or agent for execution of a transaction. They are also vulnerable to (moral hazard), the possibility that firms or agents will put their own interests or those of another customer above those of the customer or even engage in fraud. In short, unsophisticated consumers are vulnerable to incompetence, negligence and fraud.

In order to ease these asymmetric information problems, regulators often establish "fit and proper tests" for financial firms to affirm their quality (ex ante). And (ex post), it is hoped that strict enforcement of conduct of business rules with civil and criminal sanctions will deter firms from exploiting asymmetric information vis-à-vis customers. Strict enforcement of conduct of business rules also provides firms with incentives to adopt administrative procedures that ensure consumers are competently and honestly served and that employees will behave in a way that upholds the firms' reputation. Conflict of interest rules and customer suitability requirements serve a similar function.

The provision of insurance is another response to the asymmetric information problem faced by unsophisticated consumers. One of the rationales for deposit insurance is to protect unsophisticated depositors of modest means who would find it excessively costly to monitor their bank. This is articulated particularly clearly in the Deposit Insurance Directive of the European Union. Other kinds of financial contracts are also insured for the protection of unsophisticated consumers. In the United States, for example, the Pension Benefit Guaranty Corporation, a government-sponsored entity insures pension coverage up to $30,000 a year for each worker.

Disclosure requirements also help ameliorate the asymmetric information problem. Investors are often at an informational disadvantage with respect to issuers of securities. Although institutional

investors have the leverage to compel an issuer to disclose relevant data and the expertise to evaluate such data, unsophisticated consumers lack both the leverage and the expertise. For this reason governments have found it useful to standardize accounting practices, require the regular disclosure of data relevant to a firm's financial prospects and encourage the development of rating agencies, which enable even small investors to take advantage of economies of scale in gathering and analyzing data.

Disclosure concerns also extend to the way in which information is made available to the public. The United States has prohibited insider trading to ensure that corporate officials and owners with better information about the financial prospects of their companies cannot profit at the expense of non-insiders. Until recently, insider trading was not illegal in Germany nor effectively policed in Japan. But with the adoption of the Insider Trading Directive of the European Union and the disclosure of significant insider trading in Japan in the early 1990s this has changed (Herring and Litan 1995).

Reserve requirements, capital requirements and liquidity requirements designed to ensure that a financial services firm will be able to honor its liabilities to its customers, have a consumer protection (and microprudential) rationale as well as a macroprudential rationale to safeguard the system against systemic risk. In effect, regulators serve a monitoring function on behalf of unsophisticated customers of modest means.

2.3 Enhancing efficiency

Competition policy and anti-trust enforcement are the key tools for enhancing the efficiency of the financial system as can be seen in the Efficiency Enhancement column of Figure 1. In addition to prosecuting price-fixing arrangements, the main emphasis here is to minimize barriers to entry into the financial services industry. In this light, "fit and proper" tests established for consumer protection purposes appear to be anti-competitive and unnecessary. After all, the expectation of repetitive transactions with a client will give firms reason to be concerned with their reputations. This will reduce the risks of adverse selection and moral hazard to customers, except when the expected gain from taking advantage of a client is very large or when the interests of a firm's employees differ from those of the owners.

However, primary reliance on a firm's concern for its reputation is not an entirely satisfactory solution to the problem of asymmetric information. Since it takes time to build a reputation for honest dealing,

primary reliance on reputation to establish the quality of financial firms tends to restrict entry. This may result in higher transactions costs than would prevail in a perfectly competitive market. For this reason establishing ”fit and proper tests” that enable new entrants to affirm their quality ex ante may ease entry and enhance competition, although if entry hurdles are set too high, they will surely compromise efficiency objectives.

The efficient operation of the financial markets depends critically on confidence that financial markets and institutions operate according to rules and procedures that are fair, transparent and place the interests of customers first. This confidence is a public good. It increases flows through financial markets and the effectiveness with which financial markets allocate resources across time and space. But this public good may be underproduced, because the private returns to firms that adhere to strict codes of conduct are likely to be less than the social returns. Unethical firms may be able to free ride on the reputation established by ethical firms and take advantage of the relative ignorance of clients in order to boost profits. The primary efficiency rationale for conduct of business rules and conflict of interest rules is to correct this perverse incentive.

Finally, financial markets provide critical information that helps to coordinate decentralized decisions throughout the economy.[5] Prices in financial markets are used by households in allocating income between savings and consumption and in allocating their stock of wealth. These prices also help firms decide which investment projects to select and how they should be financed. Financial markets will provide better price signals and allocate resources more efficiently the better the access of participants to high quality information on a timely basis. This applies not only to information regarding issuers of financial instruments, but also to financial institutions themselves and the products they sell. Disclosure standards thus also serve an efficiency rationale as well as a consumer protection rationale.

Efficiency would also be enhanced if regulators were required to justify each new regulation with a careful assessment of its costs and benefits. This requirement is an obligation of Britain’s new financial services authority. It should be a fundamental part of the regulatory process everywhere.

[5] See Santomero and Babbel (1997) Chapters 1 and 2.

2.4 Achieving other social objectives

Governments are often tempted to exploit the central role played by the financial sector in modern economies in order to achieve other social purposes. Budget constrained governments frequently use the banking system as a source of off-budget finance to fund initiatives for which they chose not to raise taxes or borrow. Over time this politically connected lending can have a devastating impact on the efficiency and safety and soundness of the financial system as we have learned from the experience of many central and eastern European countries and the recent Asian banking crises.[6]

The housing sector is often favored by government intervention in the financial system. For example, the United States has chartered financial institutions with special regulatory privileges that specialize in housing finance. It has also promoted home ownership by extending implicit government guarantees to securities backed by housing mortgages and by allowing homeowners to deduct mortgage interest on their income taxes. In addition, until its interest rate ceilings were eliminated, the United States favored housing lenders by allowing them to pay their depositors a slightly higher interest rate than banks could pay their depositors, a policy that had the effect of enhancing the funds made available to finance housing.

Governments also channel credit to favored uses in other ways. Most countries subsidize financing for exports, sometimes through special guarantees or insurance or through special discount facilities at the central bank. Many countries also require their financial institutions to lend to certain regions or sectors. Since the enactment of the Community Reinvestment Act in 1977, the United States has required its commercial banks and thrift institutions to serve the credit needs of low-income areas.

The United States has also used regulation to achieve the social objective, first articulated by Thomas Jefferson, of preventing large concentrations of political and economic power within the financial sector, especially among banks. Until recently, the United States had restricted the ability of banking organizations to expand across state lines. Restrictions continue against bank participation in nonbanking activities.

Finally, many members of the Organization for Economic Cooperation and Development have imposed reporting requirements on banks and some other financial institutions in an effort to combat money

[6] See Santomero (1997b, 1998) for a fuller discussion of this issue.

laundering associated with the drug trade and organized crime. In the United States banks are required to report all currency transactions of $10,000 or more. Currently, Congress is considering even more stringent reporting requirements that have raised serious concerns about violations of privacy rights. Similarly the new Financial Services Authority in the United Kingdom (Davies 1998, p. 2) has adopted the objective of ”preventing … financial businesses being used for the purposes of financial crime”.

3 Why banks have been especially important

The preceding survey of the objectives of financial regulation has identified three categories of rationales that apply not only to the financial sector but also to some non-financial products and services as well. Although the means of regulatory intervention may vary from sector to sector, the objective of protecting consumers from opportunistic behavior by vendors or agents applies equally to medical services, food and many other consumer purchases. Similarly, the objective of enhancing the efficiency of markets motivates regulation in a broad range of industries in addition to the financial services industry. And, budget-constrained governments are always eager to exploit opportunities to advance broad social objectives through off-balance sheet means. Because of its status as a heavily regulated industry, the financial services industry is highly vulnerable to such attempts, but it is not unique in this regard.

However, one motive for financial regulation is distinctive to the financial services industry. Systemic risk motivates a considerable amount of financial regulation but does not apply to regulation in other industries. Moreover, within the financial sector concerns about systemic risk tend to focus on banks. Why are banks especially associated with systemic risk? What's special about banks?

Many of the products and services provided by contemporary banks are indistinguishable from products and services provided by other kinds of financial institutions. To that extent banks are less special than they once were, a topic we will investigate in section 5. However, the argument that banks are special is based on: the distinctive functions they have performed, the importance of those functions to the economy, and the consequences these functions have had for the vulnerability of their balance sheets to liquidity shocks.

First and foremost, banks have been the principal source of non-market finance to the economy. Banks gather and assess information about prospective borrowers and their investment opportunities. Using specialized human capital and financial technologies[7] they screen

[7] For a fully developed model of this function, the reader is referred to Diamond (1984), Santomero (1984) and Bhattacharya and Thakor (1993).

borrowers to identify wealth-enhancing projects that they will then finance. This may, in fact, be their most important contribution to economic performance.[8] The assets that banks acquire in this process are frequently illiquid and difficult for external parties to value without substantial effort.[9] After originating loans, banks have traditionally funded and serviced the loans, monitored the borrowers' performance and provided workout services when necessary. These efforts enhance returns from the investment project, as borrowers respond to on-going monitoring by increasing effort and by making operating decisions that adhere to the proposed purpose of the loan.[10] The bank role as monitor improves the financial performance of the project and the returns accruing to the intermediary itself.

On the liability side of their balance sheets banks mobilize savings to fund the loans they originate. The second distinctive function performed by banks is to serve as the principal repository for liquidity in the economy. Banks attract demand deposits by offering safe and reliable payments services and a relatively capital-certain return on investment. Banks have developed the capacity to mobilize idle transactions balances to fund investments while at the same time clearing and settling payments on behalf of their depositors. By pooling the transactions balances of many different transactors they can acquire large, diversified portfolios of direct claims on borrowers which enable them to meet liquidity demands while still holding substantial amounts of illiquid assets. For the economy as a whole, the smooth and reliable functioning of the resulting payments system is critical to the health of the economy.[11]

In addition to providing sight deposits, banks offer longer-term deposits that must compete directly with other instruments available in the financial markets.[12] The return on deposits must be sufficient to compensate for the risk and delayed consumption associated with accepting deposit claims on the bank.

These functions – making loans, clearing and settling payment transactions, and issuing deposits – are performed more or less simultaneously. Banks transform the longer-term, risky, illiquid claims that borrowers prefer to issue into safer, shorter-term, more liquid

[8] For a fuller discussion of this role and its effect on the economy, see Herring and Santomero (1991).
[9] For a discussion of this issue, see Gorton and Pennacchi (1990) and Santomero and Trester (1997).
[10] See Allen and Gale (1988) for a discussion of the importance of monitoring to project outcomes.
[11] Goodfriend (1989) and Flannery (1998) make this case quite effectively.
[12] This point is made theoretically and empirically in Fama (1985).

demand and savings deposits that savers prefer to hold. This asset transformation often involves maturity transformation as well. The consequence of the simultaneous performance of these three functions is that banks have balance sheets that are vulnerable to liquidity shocks. While these functions are usually mutually compatible – indeed, some researchers have argued that banks have an advantage in monitoring loans because they can observe the cash flows of their borrowers through transactions accounts (Black 1975, Fama 1985, and Lewis 1991) – a sudden, unanticipated withdrawal of the deposits that fund longer-term, illiquid loans can give rise to instability.[13,14]

Instability in the banking system can undermine confidence in the financial system and disrupt its role in facilitating the efficient allocation of resources that enhances economic growth. Moreover, it can impose massive costs on society.

From 1980 to 1995 more than three-quarters of the members of the International Monetary Fund experienced serious and costly banking problems. In 69 of these countries losses exhausted the net worth of the entire banking system, in several cases driving it to negative levels. Ten countries spent more than 10 percent of their GDP in bailing out their banking systems (Davies 1998). These direct costs of recapitalizing the banking system do not include the heavy costs imposed on the real economy due to the disruption of the payment system, the interruption of credit flows to bank-dependent borrowers, and the withdrawal of savings from the financial system.

The systemic risk rationale for the prudential regulation and supervision of banks starts from the presumption that the three basic functions that make banking special – loan origination, provision of payment services and deposit issuance – are central to the functioning of the financial system and the real economy, but give rise to bank financial structures that are vulnerable to crises. The opportunity for depositors to run from a bank arises from the fact that deposits must be redeemed at face value on short notice or demand. The motive for a bank run can arise because banks are highly leveraged – with an equity-to-asset ratio that is lower than other financial and non-financial firms – and hold portfolios of illiquid assets that are difficult to value. A rumor that a bank has sustained losses that are large relative to its equity may be sufficient to precipitate a run. Moreover, because forced liquidation of

[13] The classic references here are Diamond and Dybvig (1983), and Gorton (1988).

[14] See Kareken and Wallace (1978), Jacklin (1987), and Santomero (1991) for a fuller discussion of these issues.

illiquid bank assets can cause additional losses, once a run has begun it tends to be self-reinforcing. Even depositors who were not alarmed about the original rumor of losses may join the run once it has begun because they know that the run itself can cause substantial losses that may jeopardize the bank's solvency.

The failure of a nonbank firm is usually not a source of public policy concern in most countries.[15] Indeed, the failure of one nonbank firm often improves business prospects for the remaining firms in the industry. In contrast, a shock that damages one bank seriously can spread to other banks. Contagious transmission of shocks may occur because of actual direct exposures to the original shock and/or the failed bank or, more insidiously, because of suspected exposures. In the absence of clear and convincing evidence to the contrary, depositors are likely to suspect that the banks least able to withstand a shock have been damaged. They will attempt to protect themselves by liquidating their deposits at the suspected, weaker banks and reallocating their portfolios in favor of deposit claims on banks perceived to be stronger or claims on the government.[16] The result is a flight to quality and a banking panic that destroys not only the specific capital of the banks under pressure, but also diminishes the capacity of the financial sector to fund economically viable projects and monitor them to a satisfactory conclusion.[17]

When banks fail and markets seize up, they cannot perform their essential function of channeling funds to those offering the most productive investment opportunities. Some firms may lose access to credit. Investment spending may suffer in both quality and quantity. Indeed, if the damage affects the payments system, the shock may also dampen consumption directly. The fear of such an outcome is what motivates policymakers to act.

Prudential regulation and supervision to safeguard against systemic risk arises in the first instance from this externality. While bank managers and shareholders of a bank have appropriate incentives to take account of losses to themselves if their bank should fail – destroyed shareholder value, lost jobs and damaged reputations – they do not have adequate incentives to take account of the potential external costs to other banks and the real economy. Thus they may take riskier positions

[15] Nevertheless, the failure of very large firm tends to attract governmental attention in most countries because of its impact on employment.

[16] If depositors withdraw their balances and hold them as cash, bank reserves will contract unless the monetary authority neutralizes the shift. This may be an additional source of contagion.

[17] See the work of Bernanke and Gertler (1989, 1990) for two similar models of this phenomenon.

than if they were charged a fair market price for such risks. Prudential regulation and supervision is designed to counteract the incentive for excessive risk-taking.

4 Prudential regulation and supervision: the financial safety net

The financial safety net is an elaborate set of institutional mechanisms rigged to safeguard the economy from systemic risk that might result from contagious bank runs. This safety net can be viewed as a series of circuit breakers designed to prevent a shock to one bank from spreading through the system to damage the rest of the financial grid. For our purposes the safety net can be seen as consisting of six circuit breakers that are triggered at various states in the evolution of a banking crisis.[18]

First, the chartering function seeks to screen out imprudent, incompetent or dishonest bank owners and managers who would take on excessive insolvency exposure. This usually involves fit and proper tests that bank owners and managers must pass to qualify for a banking license. In the aftermath of the collapse of the Bank for Credit and Commerce International, which was engaged in fraud on an international scale, a number of countries established additional tests for continuance of a banking license for foreign banks.

Second, in the event that some financial institution managers do attempt to expose their institutions to excessive insolvency exposure, the prudential supervisory function seeks to prevent it. Prudential supervision is concerned both with leverage and asset quality. Capital adequacy standards, which have been partially harmonized internationally, attempt to constrain leverage risk and ensure that the bank has an adequate buffer against unanticipated losses. Supervisors attempt to control asset risk by risk-weighting capital requirements, diversification rules, restrictions on connected lending or outright prohibitions on certain kinds of assets. Bank examinations focus not only on the bank's own processes and procedures to control asset risk, but on individual bank assets to make sure that they are stated at fair value and that reserves for loan losses are appropriate.

Third, in the event that prudential supervision does not prevent excessive insolvency exposure and a damaging shock occurs, the termination authority attempts to make a regulatory disposition of the bank before it exhausts its net worth and causes losses to depositors. If

[18] This safety net is discussed in greater detail in Guttentag and Herring (1989) and Herring and Santomero (1991).

depositors could rely on prompt termination[19] before a bank's equity is exhausted, there would be no incentive to run. But the supervisory authorities face technical and political difficulties in implementing the termination function with such precision. The result is that insolvent banks are often permitted to operate long past the point at which they have exhausted their net worth.

Fourth, if the termination authority acts too late to prevent the bank from exhausting its net worth, deposit insurance may protect depositors from loss and remove the incentive for depositors to run from other banks thought to be in jeopardy. In response to the banking crisis of the Great Depression, the United States established the Federal Deposit Insurance Corporation in 1933 to provide insurance against loss for owners of small deposits. Although most other countries have long had systems of implicit deposit insurance, it is only within the last thirty years that other countries have established similar systems of explicit deposit insurance. Although deposit insurance is motivated by concerns for consumer protection as discussed earlier in section 2.2., it may also play an important role in stabilizing the banking system against shocks. The protection is imperfect, however. Even in the US, where the link to financial stability has been most explicit, deposit insurance has been limited, leaving some depositors vulnerable to loss. Thus, the possibility of a run continues.

Fifth, even if runs occur at other institutions, the lender of last resort may enable solvent institutions to meet the claims of liability holders by borrowing against assets rather than selling illiquid assets at firesale prices. Henry Thornton and Walter Bagehot articulated the rationale for the lender of last resort function during the 19th century. Usually the central bank functions as the lender of last resort because it has the resources to intervene credibly to meet any extraordinary demand for domestic liquidity. Although the members of the European Monetary Union have agreed on the powers of the European Central Bank for the conduct of monetary policy, they have not yet agreed on how – or whether – to provide lender-of-last-resort assistance to banks in the euro zone.

Sixth, even if the lender of last resort does not lend to solvent but illiquid banks, the monetary authority may protect the system from cumulative collapse by neutralizing any shift in the public's demand for

[19] The "termination" of a bank means that the authorities have ended control of the bank by the existing management. Termination may involve merging the bank with another, liquidating it, operating it under new management acceptable to the authorities or some combination of these actions.

cash thus protecting the volume of bank reserves. In this way the monetary authority can prevent any flight to cash from tightening liquidity in the rest of the system. This is precisely what the US monetary authorities failed to do during the Great Depression. But the lesson was not wasted. Most modern monetary authorities are committed to maintaining policy control over the reserve base.

In the major industrialized countries, the various circuit breakers that comprise the financial safety net have been generally successful in preventing a problem at one institution from damaging the system as a whole. In the United States, for example, the safety net which was constructed in the 1930s has virtually eliminated the contagious transmission of shocks from one depository institution to the rest of the system. Similarly in the recent Swedish banking crisis, the Riksbank succeeded in preventing a contagious transmission of shocks to the rest of the financial system and minimized the damage to the real economy.

In effect, banking systems in most market economies operate with the implicit support of their regulatory authorities. With the possible exception of New Zealand, where the authorities have explicitly taken down their safety net for banks,[20] the intervention of the regulatory authorities in time of crisis is rationally expected in every market economy. Financial safety nets have reduced the frequency of bank runs, banking panics, and financial disruption. However, these safety nets may have worked too well. Depositors and other creditors have come to rely on their bank's access to the safety net as a protection against loss with the consequence that they exercise only limited surveillance over riskiness. The pricing of bank liabilities depends heavily on the bank's presumed access to the safety net. The result is that banks are not penalized for taking greater risks as heavily as they would be if they did not have access to the safety net.[21] Consequently, banks take on greater risks.[22]

This moral hazard feature of the safety net has contributed to the frequency and severity of banking problems, which appear to be rising. In both Eastern Europe and the Far East we have ample evidence of institutions that have assumed excessive risk and suffered severe consequences. As noted above, from 1980 to 1995 three-quarters of the members of the IMF experienced serious and costly problems. For

[20] New Zealand's policy is especially credible because all major banks are owned by foreign residents.

[21] There are a large number of empirical studies on this point. See Gorton and Santomero (1990), Ellis and Flannery (1992) and Flannery and Sorescu (1996).

[22] For empirical evidence see Keeley and Furlong (1987, 1991).

example, the real cost of the Savings and Loan crisis in the US has been estimated at less than five percent of GDP, and current estimates for the Japanese economy center are five to ten times this proportion. In less developed economies, where the magnitude of the crisis is even greater and fewer resources are available for resolution, the costs associated with the financial safety net have exceeded the country's financial capacity.

This has led many to argue that financial regulation and the safety net itself needs some adjustments. Indeed, perhaps the entire approach to regulation needs to be reexamined to find a better way to obtain the benefits associated with a well functioning financial sector, but at a lower cost.

5 Optimal regulation in the static case: pricing risk to counter moral hazard

Since the safety net distorts incentives for risk-taking by insulating institutions and their creditors from the full consequences of their risky choices, and the consequences are seen as quite costly, the challenge for optimal regulation is to increase market discipline. In principle, this may be accomplished in a number of ways–risk-rated deposit insurance premiums, least-cost resolution combined with prompt corrective action, a subordinated debt requirement or a narrow bank structure. In practice, none of these remedies is entirely satisfactory.

5.1 Risk-rated deposit insurance premiums

Ideally, the deposit insurer could set risk premiums for deposit insurance that would be identical to the premiums that depositors would demand if the safety net did not exist. In the US, the Federal Deposit Insurance Corporation Improvement Act of 1991 (FDICIA) required that the Federal Deposit Insurance Corporation (FDIC) implement a system of risk-rated deposit insurance premiums. However, to date the result has been very crude. The maximum price difference between the safest and the most risky bank when the system was implemented was 8 basis points. This differential was far below the differential that would be charged in debt markets for such large differences in risk.[23] It is also far less than the differences in actuarially fair insurance premiums estimated from option pricing models.[24]

Although the FDIC's approach was especially crude, it is difficult to see how the ideal system could be implemented effectively. The deposit insurer faces two problems. First, the deposit insurer must be able to measure the bank's current net worth, evaluate its risk exposure, and

[23] For example, the differential between B-rated and AAA-rated bonds is typically well over 100 basis points.

[24] Kuester and O'Brien (1990), for example, estimated that fair premiums for most firms would be very low, less than 1 basis point, while a few very risky banks had fair premiums in the 1000s of basis points.

assess how the bank's net worth will vary under alternative scenarios. Such information is not currently available to the regulators and in view of the opacity of most banks, it would be very costly to obtain and verify. Second, the deposit insurer must be able to constrain the ability of the insured bank from increasing its exposure to risk after the deposit premium is set. This would require an ex post adjustment procedure to constrain moral hazard that has yet to be satisfactorily specified.[25]

5.2 Prompt corrective action and least cost resolution

FDICIA implemented yet another market-mimicking approach to countering the moral hazard incentive implicit in the safety net. The aim was to make sure that banks would not be able to operate without substantial amounts of shareholders' funds at risk.[26] It attempted to reduce the scope for forbearance by replacing supervisory discretion with rules that would mimic the conditions that banks impose on their own borrowers when their financial condition deteriorates.[27]

The FDICIA rules are designed to stimulate prompt corrective action as soon as a bank's capital position deteriorates. The regulatory sanctions become increasingly severe as a bank's capital position declines from the well-capitalized zone down through three other zones to the critically undercapitalized zone in which the supervisor must appoint a receiver or conservator within 90 days. The objective is to provide the bank's owners with incentives to take prompt corrective action by recapitalizing the bank or reducing its risk exposures before its capital is depleted. This is a strategy of deploying the termination authority in a way that substitutes for market discipline.

[25] Some researchers have argued that private insurance companies should provide some deposit insurance coverage. But private insurers would face the same challenges that the government insurer faces. Moreover, if the government continues to be concerned about systemic risk, its problem may shift from one of guaranteeing banks to guaranteeing private insurers of banks.

[26] One of the clear lessons from the S&L debacle in the United States is that losses surge as institutions become decapitalized and shareholders and managers are tempted to gamble for redemption.

[27] The fundamental analysis underlying this approach to bank regulation may be found in Benston and Kaufman (1988) and Benston et al (1989).

FDICIA also attempted to end two other sources of distortion implicit in the safety net. The United States, like many other countries, has provided implicit deposit insurance for all depositors at large banks. This subsidy has been provided in two different ways. First is the practice of using purchase and assumption transactions in which the institution purchasing the assets of a failing institution assumes all of its liabilities. FDICIA reduced the scope for these transactions by requiring that the FDIC use the least costly method of resolution under the assumption that its only liability is for explicitly insured deposits.

Second is the practice of extending lender-of-last resort assistance to insolvent banks. This provides uninsured depositors the time and opportunity to flee before the bank is closed. FDICIA attempted to deter such practices by depriving the central bank of the protection of collateral for advances extended to banks near insolvency. There is a major exception if the Fed and the Secretary of the Treasury agree that such advances are necessary to prevent ”a severe adverse effect on … the national economy”. Whether this will be a significant constraint on Fed behavior when a large bank is in jeopardy remains to be seen. But there is at least some reason to doubt that protection will be automatic and this should enhance market discipline.

FDICIA’s prompt corrective action measures are subject to the same problems as risk-rated deposit insurance. Both depend on accurate measurement of the economic value of a bank’s capital position and its potential risk exposure. At a minimum this would require adoption of a mark-to-market accounting system.[28] Moreover, capital adequacy will need to be monitored in shorter intervals than in the past since a bank active in derivatives markets can change its risk exposures drastically within a very short period.

5.3 Subordinated debt

A rule that banks fulfill a specified part of their capital requirements with subordinated debt provides an alternative way to increase market discipline on banks. Subordinated debt is junior to all claims other than equity and so serves as a buffer against losses by the deposit insurer. Subordinated debt has some of the characteristics of ”patient money” because it typically has a maturity greater than one year and cannot be redeemed quickly during a crisis. Subordinated creditors have strong

[28] FDICIA called for accounting reforms that would move regulatory measures of capital closer to actual market values, but no real progress has been made.

incentives to monitor bank risk-taking and impose discipline – provided that they believe that they will not be protected by the safety net in the event of failure. Indeed, their loss exposure is similar to that of the deposit insurer. They are exposed to all downside risk that exceeds shareholders' equity, but their potential upside gains are contractually limited. In contrast to shareholders that may choose higher points on the risk-return frontier, subordinated creditors (like the deposit insurer) generally prefer safer portfolios and are likely to penalize banks that take significant risks.

The price discipline of traded subordinated debt – which is actively traded in secondary markets – is a much quicker and perhaps more precise way of controlling bank risk taking than regulatory measures which are often blunt and cumbersome to deploy. A falling price of subordinated debt can alert other creditors about the condition of the bank or actions of the managers, creating a broader market reaction. Moreover, market prices are more forward looking than regulatory examinations and may provide regulators with valuable information on the market's perception of the risk taken by banks (Horvitz 1983).

When bank risk increases unexpectedly, banks may not have to pay higher rates or face possible quantity discipline until their subordinated debt matures. For this reason, subordinated debt proposals generally require that banks stagger the maturities of their subordinated debt so that a modest proportion matures each quarter. In this way market discipline – through price and quantity sanctions – may be effective and informative, but sufficiently limited in magnitude to provide time for crisis resolution or orderly termination.

Critics of subordinated debt requirements emphasize that subordinated debt holders would face the same informational asymmetry problems that the deposit insurer faces, but without the authority to conduct detailed examinations.[29] They also question whether secondary markets in subordinated debt would be broad and deep enough to provide reliable price signals.

[29] While disclosure practices are endogenously determined, one might expect subordinated debt holders to demand fuller disclosure. As Kane (1995, p. 455) observes "an outside risk sharer must be able to persuade institutional managers to open their books in ever-changing and nonstandard ways".

5.4 Narrow bank proposals

Another approach to correcting the distortion of incentives that arises from the safety net is to narrow the range of assets that the insured unit of a bank can hold so that the risk to the deposit insurer is essentially zero and so that whatever remaining subsidy inherent in the safety net does not spill out to distort other lines of business. ”Narrow bank” proposals (Litan 1987, Pierce 1991, and Miller 1995) require that insured deposits be invested only in short-term Treasury bills or close substitutes. Banks would also issue non-guaranteed financial instruments such as commercial paper to fund conventional bank loans, just as finance companies and leasing companies now do.

Alternatively, most of the benefits of the transparency and simplicity of this approach could be maintained, while allowing greater flexibility in portfolio choice, if banks were permitted to hold not only short-term Treasury bills but also other assets that are regularly traded on well-organized markets and can be marked to market daily. This could be implemented in two ways: (1) the ”secure depository” approach in which institutions would be required to form separately incorporated entities taking insured deposits and holding only permissible, marketable assets; or (2) the ”secured deposits” approach in which insured deposits secured by a lien on a pool of permissible assets would be in a corporate entity holding other assets and liabilities (Benston et al, 1989). Capital requirements for the ”secure depository” (or the analogous excess collateral requirements for ”secured deposits”) would be set to ensure that the chance of insolvency between daily mark-to-market points is reduced to some minimal probability. This would, in effect, permit the termination function to be performed with the precision necessary to protect depositors and the deposit-insuring agency from loss.

Critics argue that the narrow bank approach does not address all of the features that make banks special and especially vulnerable to systemic risk. Government might still feel compelled to exercise prudential oversight over the other parts of financial institutions that provide credit to difficult-to-monitor borrowers and issue liabilities that substitute for lower-yielding deposits in the narrow bank. The commitment to constrain the safety net to the narrow bank might not be credible and thus the distorted incentives for risk-taking may continue.

6 Looking beyond the static view: Banks have become less special

The case for prudential regulation of banks to safeguard against systemic risk rests on the argument that banks are special. This stems from their central role as providers of credit, as repositories of liquidity and as custodians of the payment system which gives them a balance sheet structure that is uniquely vulnerable to systemic risk.
Indeed, in most countries, banks retain a central role as the most important providers of credit. (See Figure 2.)

Figure 2. Banks' share in financial intermediation, 1994

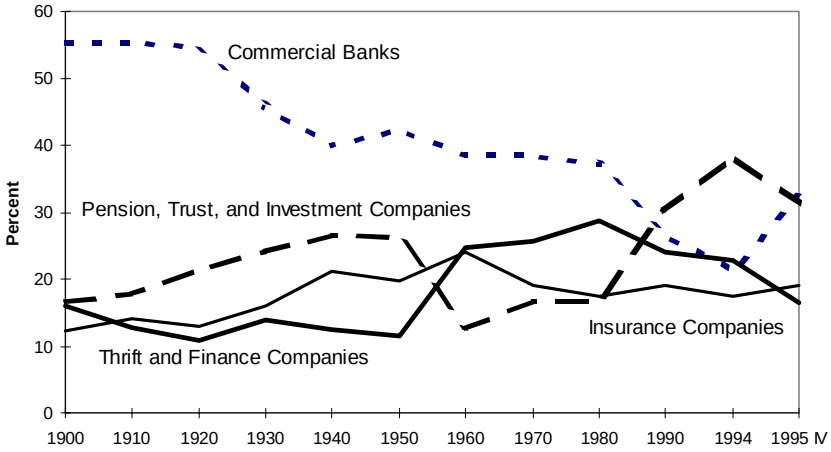
Germany	77 %
Japan	79 %
Sweden	79 %
United States	23 %

Source: Bank for International Settlements, Annual Reports and IMF International Financial Statistics.

The one exception is the United States, where banks have experienced a marked decline in their share of the assets held by the financial sector.[30] Although this declining share is often assumed to be a recent phenomenon, in fact the trend was apparent in the 1920s. Indeed, the 1920s were an era much like the last two decades in which the share of assets held by banks declined and that of pension funds trusts and investment companies grew. In the broader historical context the anomaly may have been the relative stability of the bank share of total assets from the 1940s through the mid-1970s. Figure 3 offers some evidence of this for the US case.

[30] Allen and Santomero (1997) present evidence of a trend away from bank finance in other leading countries.

Figure 3. Relative shares of total financial intermediary assets, 1900-1995 Q4

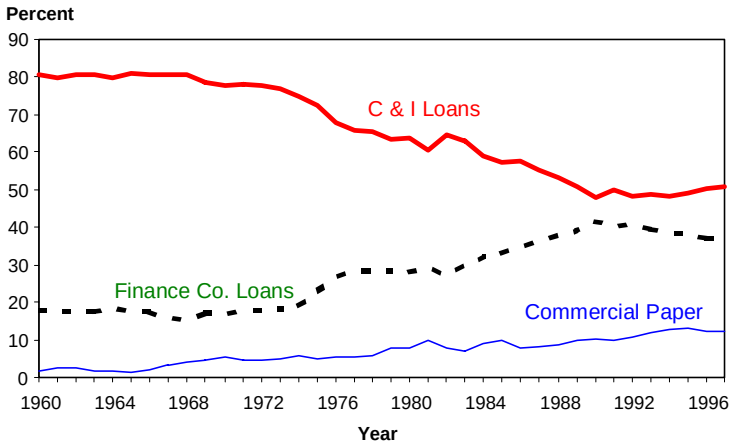


Source: Federal Reserve Bulletin, 1995.

The reasons for this long-term trend and its recent acceleration are, no doubt, numerous. However, technology is clearly an important force. Advances in technology have led to innovations in financial instruments and institutions that have blurred traditional product-line boundaries that formerly distinguished banks from other financial institutions. The ability to call up information cheaply at any time from virtually any location has enabled other financial institutions to design new products that compete effectively in terms of price and quality with traditional bank products. Regulators have generally responded to these developments by liberalizing some of the regulatory restrictions that constrained competition among banks and between banks and other financial institutions including foreign financial institutions.

The impact has been most dramatic on the asset side of banks' balance sheets. The increased institutionalization of consumer savings, especially in pension plans, life insurance and mutual funds, has given other institutions the scale to assess and diversify credit risk in competition with banks. Improved disclosure standards have made information regarding the creditworthiness of borrowers, which was once the proprietary domain of bankers, publicly available. Credit-rating agencies have grown in importance and perform the kind of analysis that was once the comparative advantage of banks. Moreover, when credit rating agencies have turned their attention to banks, they have often concluded that banks are less creditworthy than many of their prime borrowers.

Figure 4. Commercial and industrial loans as a share of short-term business finance, 1960-1998



Source: Board of Governors of the Federal Reserve System, "Flow of Funds accounts", 1998.

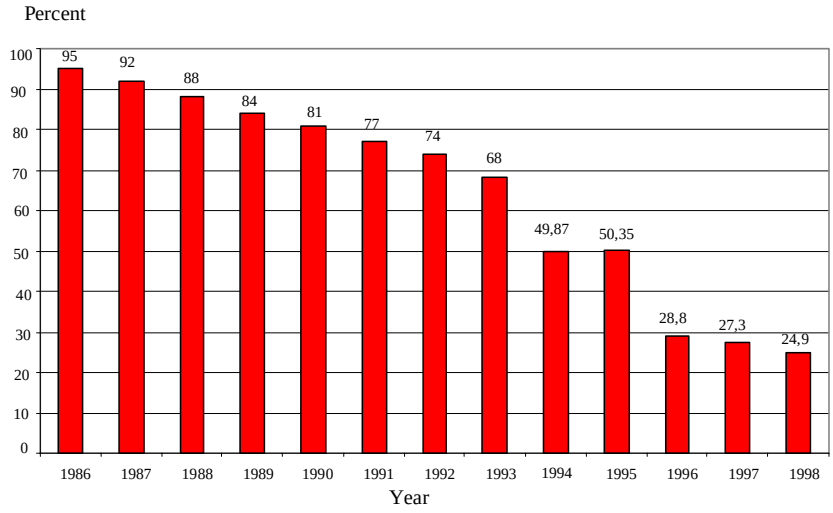
The decline in the role of banks as intermediators of credit risk has been most pronounced in a US context with regard to business finance as Figure 4 indicates. Banks have lost ground to other, less regulated intermediaries such as finance companies and to securities markets, especially the commercial paper market and the high yield securities market. Indeed, some cynical observers have asserted that the typical bank loan is simply a less liquid, under-priced junk bond.

The decline in business lending is also mirrored in consumer lending. (See Figure 5.) Banks have lost market share to nonbanks such as AT&T, GMAC, GE & Morgan Stanley Dean Witter. Twenty years ago, banks completely dominated the card-transactions processing business. Now, banks hold less than 25 % of receivables and close to 80 % of credit card transactions are processed by nonbanks such as First Data Resources.[31]

Increasingly, nonbank, single-purpose providers have successfully competed for some of the most profitable traditional bank products. The development of securitization techniques has transformed the way in which many kinds of credit transactions – which would previously have been conventional bank loans – are structured.

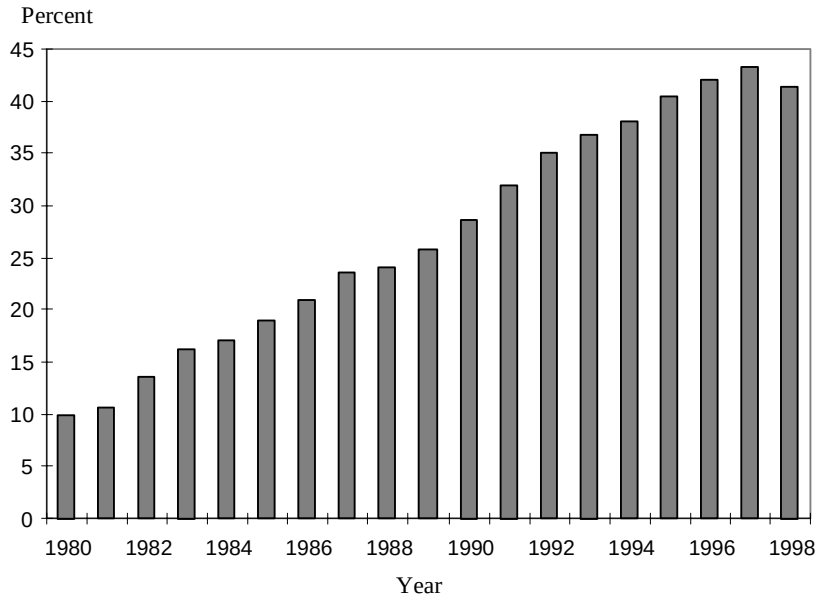
[31] See Business Week, June 12, 1995, p. 70.

Figure 5. Bank market share of credit card receivables, 1986-1998



The growing importance of securitization is especially obvious in the transformation of the traditional mortgage. (See Figure 6.) Formerly, a bank originated, funded and serviced the mortgage until it was repaid. Now one firm may originate the mortgage. Another firm may fund the mortgage or pool the mortgage with others and partition the anticipated flow of income from the pool into marketable securities that will appeal to particular groups of investors around the world. Another firm may insure the pool of mortgages to facilitate this process. The servicing of the mortgage may be allocated to yet another specialist firm that has data processing expertise. The consequence is that mortgages will be funded at lower cost than if firms were obliged to hold mortgages to maturity and what was once an illiquid bank asset is transformed into a highly marketable security. This unbundling can be executed so smoothly that the mortgagee may be entirely unaware that it has taken place. These techniques have been successfully applied to many other kinds of credit transactions including credit card receivables, auto loans, and small business loans.

Figure 6. Securitized mortgages as a percent of total mortgages, 1980-1998

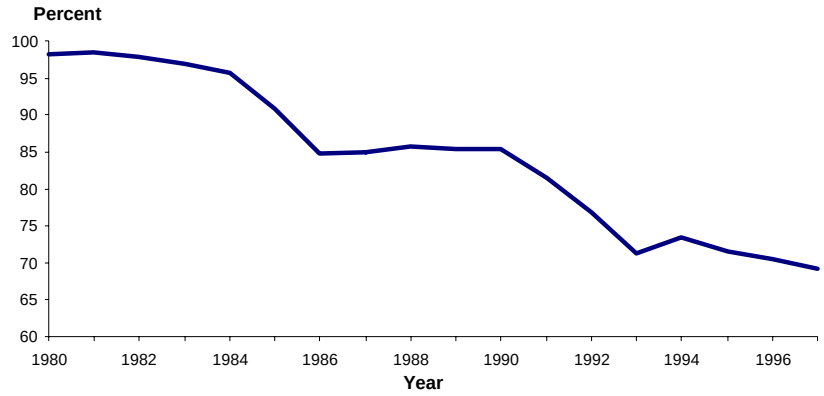


Source: Board of Governors of the Federal Reserve System, Flow of funds accounts," various years.

Banks are also losing ground on the liability side of their balance sheets. As the baby boom generation matures and inherits wealth, consumer demand will shift from credit products to savings products. This trend is apparent in most industrial countries. In the United States over the next twenty years the population under age 50 will remain the same as it is today, but the population older than 50 will double. The traditional bank entry in the competition for consumer savings – the time and savings account – is deservedly losing ground to mutual funds that have much leaner cost structures and can offer higher returns.[32] Bank time and savings deposits have declined steadily relative to fixed-income mutual funds since 1980. (See Figure 7.)

[32] See Santomero and Hoffman (1998) for even more evidence of this trend away from banking institutions.

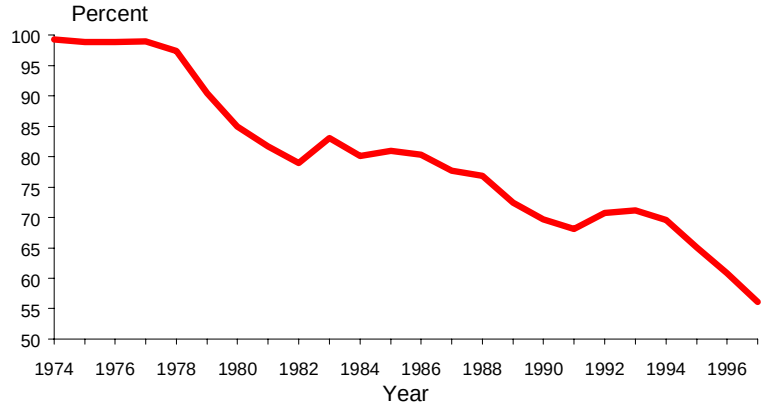
Figure 7. Bank time and savings deposits decline relative to fixed-income mutual funds, 1980-1997



Source: Investment Company Institute Mutual Fund Fact Book, 1998; and Federal Deposit Insurance Company, *Historical Statistics on Banking*, 1997.

New technology – often introduced by nonbanks – is jeopardizing even the fundamental role of banks in facilitating payments. (See Figure 8.)

Figure 8. Checkable deposits decline relative to money market mutual fund shares, 1974-1998



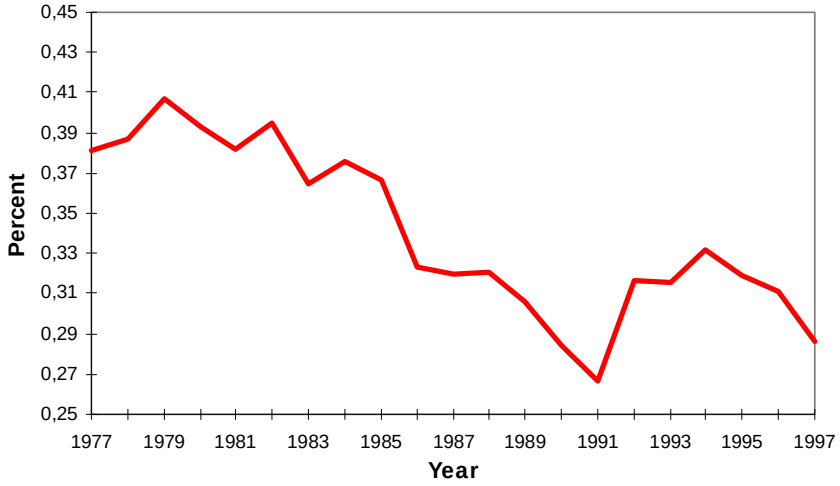
Source: Investment Company Institute Mutual Fund Fact Book, 1998; and Federal Deposit Insurance Company, *Historical Statistics on Banking*, 1997.

Many mutual fund families and most brokerage houses offer cash management accounts that permit individuals to arrange for their salaries to be automatically deposited in their cash management accounts from which routine payments can be made automatically and irregular payments may be made by phone twenty-four hours a day. Personal

checks may be drawn on the money market account. In addition, money market accounts can be linked to a credit card that also functions as a debit card at automated teller machines for cash needs. Although payments through the account are cleared through a bank, the role of the bank is a regulatory artifact, not an essential, unique part of the transaction.

Looking ahead, it is not clear how retail customers will want to deal with their banks in the future – or, indeed, whether they will want to

Figure 9. Net interest income less charge-offs as a percent of financial sector GDP 1977-1997



Source: Survey of Current Business; and Federal Deposit Insurance Company, *Historical Statistics on Banking*, 1997.

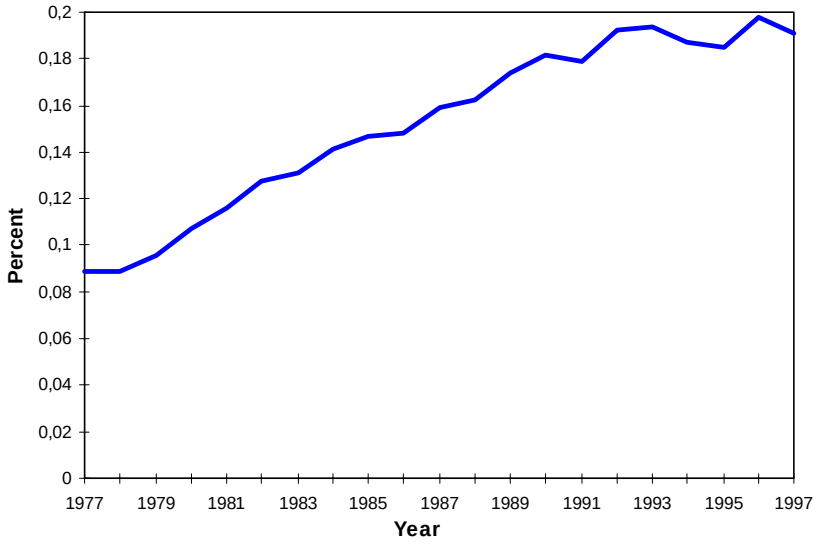
deal with banks at all. It is clear that retail customer want ubiquitous access, speed and reliability. Channels for delivery of banking services are proliferating and some by-pass banks altogether. Cyber cash or e-money is the most revolutionary concept. In principle, money can be downloaded to a personal computer or a palm-sized electronic wallet or smart card and used to make purchases over the internet or even from vendors on the street. Banks retain the advantage – due in part to deposit insurance – of consumer trust, but other firms – e.g. software, telephone or cable companies – may have advantages that will prove to be more potent in the world of cyber cash.

In view of the declining role of the traditional intermediation business, it is not surprising to see the importance of net interest income to both the banking sector and the economy as a whole has fallen in the

US. (See Figure 9.) Because this decline in the intermediation business is economically motivated and technologically driven, it is likely to be both irreversible, and global in impact.

Although the intermediation business has declined, banks have managed to prosper nonetheless by shifting from traditional intermediation functions to fee producing activities such as the trusts, annuities, mutual funds, mortgage banking, insurance brokerage and

Figure 10. Noninterest income as a percent of financial sector GDP 1977-1997



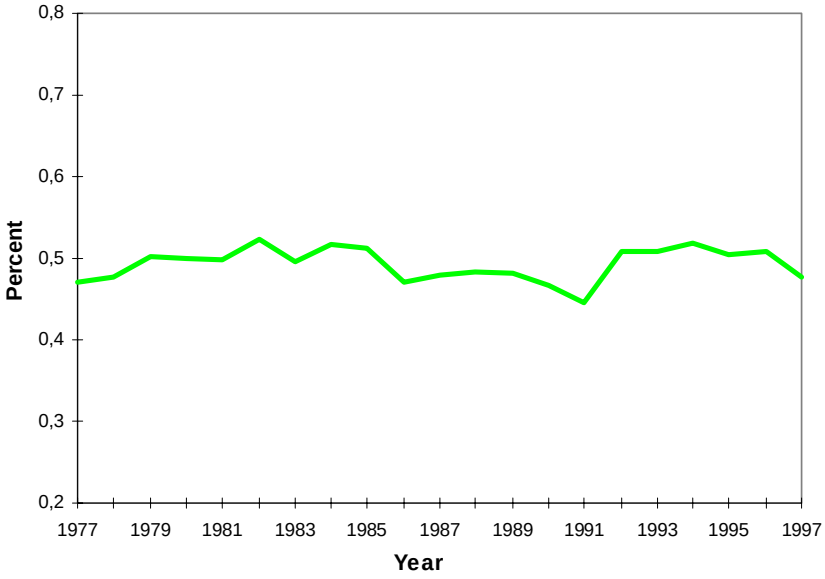
Source: Survey of Current Business; and Federal Deposit Insurance Company, *Historical Statistics on Banking*, 1997.

transactions services. (See Figure 10.) Notwithstanding the constraints on allowable bank activities in the US, imposed by the Glass-Steagall Act and the Bank Holding Company Act, banks have managed to develop new lines of business to compensate for the decline in the traditional intermediation business.

Overall, banks are holding their own, (see Figure 11), but with a very different configuration of earnings.[33] Spread income accounted for about 80 % of bank earnings only a decade ago. Now most large regional and money center banks earn more than half their income from fees and trading income.

[33] Boyd and Gertler (1994) emphasized this point.

Figure 11. Bank value added as a percent of financial sector GDP, 1977-1997



Source: Survey of Current Business ; and Federal Deposit Insurance Company, *Historical Statistics on Banking*, 1997.

The result is that banks are markedly less special in the United States than they were even a decade ago. They are no longer the primary source of business and consumer finance. Neither are they the main repository of liquid savings for the financial system. They do remain custodians of the payment system and for that reason concerns about systemic risk persist. The principal source of concern is what Flannery (1998) has described as ”credit-based” mechanisms for the exchange of large-value payments. The problem is that many (but not all)[34] national payment systems permit banks to run substantial overdrafts in the process of clearing and settling payments. In effect, the systems rely on the equity of participating banks to control default risks and, failing that, the willingness of governments to intervene and support the system in the event of crisis.

The G-10 Committee on Payment and Settlement Systems has attempted to measure and quantify exposures that result from settling foreign exchange transactions. The Allsopp Report (BIS, 1996) concluded that exposures could exceed three days worth of trades with

[34] Australia, New Zealand, Switzerland and the new TARGET system for clearing and settling euro payments operate without permitting participating banks to run overdrafts.

exposures to a single counterparty in excess of a bank's capital. The failure of a counterparty could set off a chain reaction that might bring the whole system to a halt.

This kind of credit exposure is especially insidious. Although it is relatively easy to measure and monitor direct bi-lateral exposures to a particular bank, it is virtually impossible to evaluate indirect exposures. Humphrey (1986) illustrated this point when he simulated the consequence of the failure of a single settling participant in the Clearing House Interbank Payments System (CHIPS) system in the era before bi-lateral credit limits, net debit caps and collateralization arrangements were established. He found that the failure had devastating knock-on effects to many other banks in the system as the original default caused other banks to default which caused still more banks to default. When Humphrey tried the simulation on another day during the same month, the scope of the devastation to the payments system was comparable, but a different set of banks was effected. These indirect exposures are opaque not only to outsiders monitoring the banks, but also to the banks themselves.

Under pressure from the regulatory authorities, led by the G-10 Committee on Payment and Settlement systems, private sector clearing houses and central banks have been taking measures to reduce and eventually eliminate overdrafts. Real-time gross settlement, in which settlement is made payment for payment without overdrafts, is the objective. Indeed, there are plans for implementation of a Continuously Linked Settlement Bank to eliminate default risk from the clearing and settlement of foreign exchange transactions. Collateralization techniques have long been used to eliminate default risk from the settlement of futures contracts and they have also been used to eliminate the risks that Humphrey illustrated in the CHIPS system. The private sector, following proposals by the Group of Thirty (Global Derivatives Study Group, 1993), has pressed for strengthening the legal infrastructure to support netting of gross exposures so that smaller, net amounts, need to be settled.

In support of these efforts to reduce credit risk in the payments systems central banks in the three largest economic regions have committed to expanding their hours of operation so that payment against payment transactions can take place in bank reserves. Since December 1997, the Federal Reserve has extended the operating hours of Fedwire from 12:30 a.m. to 6:30 p.m. Eastern Standard Time so that it overlaps with the entire European business day and two-and-one-half hours with Japan. The TARGET system for settling euros began operations in January 1999 from 7:00 a.m. to 6:00 p.m. Central European Time. And, by 2001 the Bank of Japan will open its Japan Net from 9:00 a.m. to

7:00 p.m. Tokyo Time so that it will overlap Fedwire for four-and-one-half hours and TARGET for four hours.

Flannery (1998, p. 30) sees this movement away from a credit-based payments system as ”eliminating the need for prudential government supervision of large financial firms”. Once the issue of bank solvency has been divorced from the integrity of the payments system, the last remaining aspect in which banks are special will have ended. When banks are no longer a source of systemic risk, the safety net can be taken down and banks can be regulated like other financial firms.

7 ”Optimal” regulation in the transition: some simple prescriptions

Banks everywhere have been subject to intense regulatory oversight and limits, to one degree or another, on allowable activities. Banks in the United States have been subject to relatively tight activity restrictions that have, until quite recently, prevented them from entering many lines of the investment banking business or providing most kinds of insurance to their customers. Nonetheless, they and their counterparts throughout the world have managed to restructure their businesses so that they are much less dependent on traditional intermediation income than they were even a decade ago. As we have seen, most of the large American banks now earn a greater portion of their income in the form of fees and trading revenue with less from spread income.

Figure 12. Relative sources of revenue of major banks

	1986-1988			1993-1995			Percentage
	Spread	Fees	Ratio	Spread	Fees	Ratio	Change
Belgium	1.69	.47	3.60	1.28	.50	2.57	28.76 %
Denmark	2.76	.55	5.04	3.68	.66	5.55	-10.02 %
Germany	2.25	.56	4.04	2.11	.57	3.70	8.56 %
France	2.00	.48	4.17	1.26	.90	1.40	66.49 %
Iceland	5.97	1.78	3.35	4.76	2.22	2.14	36.13 %
Spain	3.90	.83	4.68	3.02	.93	3.24	30.80 %
UK	3.20	1.85	1.73	2.37	1.83	1.30	25.07 %
Sweden	2.51	1.13	2.22	2.65	2.04	1.30	41.62 %

The same trend is apparent for their counterparts throughout Europe and the major OECD nations. Figure 12 illustrates this. Using OECD data, it contrasts the ratio of interest income to fee income over two discreet periods, 1986-1988 and 1993-1995. Notice that in each case, with the exception of Denmark, the relative importance of on-balance sheet net interest income has declined over the period. (The Danish case can be explained by the volatility of Danish financial reports due to their mark-to-market accounting practices.)

European banks by tradition have long been permitted to offer a much broader range of services than their American counterparts. They have been active for some time in underwriting, the direct purchase of equity in the industrial sector, investment management, and a wide array of securities activities. In a recent study of comparative financial systems, Barth, Nolle and Rico (BNR) (1997) illustrate the wide range of bank activity across Europe and around the world. Their comparison across the G10 and other EU nations (replicated here as Figures 13, 14, and 15) illustrates that European banks have broad charters and are fully competitive across the entire range of universal banking products.

Swedish institutions tend to look quite similar to many of their European counterparts. Sweden's traumatic experience with real estate finance at the beginning of the decade has led Swedish regulators to be somewhat more restrictive with regard to real estate activities than other regulators in Europe and this is presumably the reason for BNR's designation of the Swedish regulatory regime as "somewhat restrictive". But, in general, Swedish banks may offer a wide array of permissible services and have a broad range of affiliations.

In view of the more liberal regulatory regime in Europe, it is surprising that European banks continue to be relatively heavily reliant on traditional intermediation services. Spread income is still more important to European banks than non-interest income. In this regard, European banks remain more "special" than their counterparts in the United States. Nonetheless, they are subject to the same forces of technological advance, innovations in financial instruments and institutions and heightened competition as banks in the United States. This difference is likely to disappear over time, as is evident in Figure 12 reported above.

In light of this unmistakable trend, what should be the role of financial regulation? Here the message should be clear. If, as we have argued, it is not possible to fully correct the distorted incentives for risk taking that are implicit in the safety net, it is important to facilitate and nurture the trends that will ultimately make the safety net unnecessary. If the safety net cannot be patched adequately, the best course of action may be to advance the conditions under which it may be taken down.

How can this be accomplished, or at least supported by regulating authorities? Here, we offer several simple prescriptions. First, the authorities should encourage the introduction of technological improvements that are lowering the costs of information and the costs of storing, retrieving and organizing these data. They should be active supporters of competition in the technology and communication sectors. These technical advances will intensify international financial

integration. Already, major investors routinely compare returns across a wide array of international financial arenas, and major borrowers choose from a menu that includes not only traditional domestic sources, but also numerous international alternatives.

Technical advances will accelerate the pace of innovations in financial products and institutions. The ability to call up information cheaply at any time from any location will enable institutions to design new products that will better serve the needs of their customers. This may often be a cheaper substitute for a service provided by a heavily regulated institution and thus will add to pressures to liberalize regulation where it is counterproductive. Institutions will introduce new processes and streamline existing ones. Cheap and easy access to customer data and the application of expert systems will enable financial firms to target particular market segments more efficiently and to distribute multiple financial products at very low marginal cost. Technical innovations will also enable financial firms to assess the profitability and riskiness of each line of business with greater accuracy and timeliness and thus to manage capital more efficiently. As firms employ sophisticated management information systems to determine which lines of business to expand and which to exit, new kinds of financial institutions will inevitably arise.

The *second* prescription is for regulators to resist the temptation to re-regulate or promulgate regulations that will forestall the inevitable financial restructuring that is part of this change process. The fundamental thrust of the forces of change – intensified international financial integration, increased innovations in financial instruments and institutions, and the liberalization of financial regulation – is to heighten competition in the financial services industry. Greater competition will be painful to many firms. It is likely to reduce the prices of financial services, diminish profit margins, reduce market shares both globally and locally and reduce the franchise value for some institutions. There will be strong political pressures to restrain these forces of creative destruction by providing implicit and explicit subsidies to local firms in general or selectively to firms in distress. There will also be attempts to restrict entry so as to slow the pace of change. Thus, the important challenge for regulation will be to maintain pro-competitive policies, which in the long run are in the national interest. This is not an easy task.

In addition, the regulatory authorities will be pressured to exercise forbearance to enable weak firms to adjust to new forces of competition or to support local firms facing aggressive external competition. It is important for the authorities to resist. Not only do such actions create a barrier to entry and maintain excess capacity in the market, but also they

put the deposit insurer and taxpayers at significant risk. Entrenched managers may resist competitive pressures to downsize, streamline or merge, and instead take on riskier projects to try to maintain the size and profitability of their institutions. Since a regulatory response is likely to lag a bank's actual risk exposures, it could have serious consequences on both the financial sector and the real economy that depends upon it for capital.

Inaddition, the standard competition policies will need to be reassessed. Anti-trust policy, for example, has an important role to play because incumbent firms may try to bar new entrants. However, anti-trust enforcement will need to be reconsidered because the relevant product markets may often be global and extend across a range of competitors that includes other financial institutions as well as banks.

The conflict of interest rules, and "fit and proper" entry tests should also be reexamined. Care should be taken to make sure that they are calibrated to accomplish consumer protection objectives and efficiency objectives only. It is important that they not deter new entrants unduly.

Third, since market discipline will increasingly substitute for prudential regulation, it is important to assure that both regulation and the regulatory staff are of a quality that is consistent with global standards. In terms of the former, increasing emphasis must be placed on market values throughout the regulatory process, and it is important to improve disclosure standards as well. Banks should be encouraged, if not required, to report their exposures to risk in terms of the market value of their assets, liabilities and off-balance-sheet positions. This will enable customers, creditors and shareholders to evaluate their prospects and react accordingly. They should also be required to report on the risk management and risk control systems in place. The development and use of rating agencies should be encouraged.

In terms of the latter, the quality and expertise of the regulation and examination staff must keep pace with the escalating standards of the global marketplace. In many respects the infrastructure of any regulatory regime is the people that enforce and oversee regulations that have been put in place by the political process. In this changing financial sector investments must be made in this infrastructure to insure that the regulatory staff are cognizant of global market trends and are capable of assuring the health of institutions under their regulatory mantle.

The safety net will undoubtedly be subjected to substantial new strains before it can be taken down. The transition will be painful for regulators and for entrenched firms. But, the gain will be a much stronger, more flexible financial system that serves its customers at much lower cost.

Figure 13. Permissible banking activities and bank ownership in selected EU and G-10 countries: 1995

Country and Bank Supervisor(s)	Securities[1]	Insurance[2]	Real Estate[3]
France Credit Institutions Committee, Bank Regulatory Commission, and Banking Commission	Unrestricted; conducted either directly in bank or through subsidiaries. No firewalls mandated.	Permitted; sale of insurance products/ services may be conducted directly in bank, but underwriting must be done through subsidiaries.	Permitted; either conducted directly in bank or through subsidiaries, but limited to 10% of the bank's net income.
Germany Federal Banking Supervisory Office and Deutsche Bundesbank	Unrestricted; conducted directly in bank. No firewalls mandated	Restricted; conducted as principal only through insurance subsidiaries, which are supervised by the Insurance supervisory Office. Insurance regulation does not allow any business other than insurance business being carried out by an insurance firm. However, a bank may conduct insurance activities as agent without restrictions	Permitted; Investments in equity and real estate, calculated at book value, may not exceed a bank's liable capital, but unlimited through subsidiaries.
Italy Bank of Italy	Unrestricted; conducted either directly in bank or through subsidiaries. However, for brokering and dealing in securities listed on an Italian exchange other than Italian government and government-guaranteed securities, only through a special subsidiary. Firewalls are mandated.	Permitted; sale of insurance products/services may be conducted directly in bank, but underwriting must be done through subsidiaries.	Restricted, generally limited to bank premises.
Japan Ministry of Finance (primary responsibility) and Bank of Japan	Restricted; only bonds (not equities) and only through securities subsidiaries. A bank can only own more than 50% of a securities firm with permission from the Ministry of Finance and Fair Trade Commission. Firewalls are mandated.	Prohibited	Restricted; generally limited to bank premises.
Sweden Financial Supervisory Authority	Unrestricted; conducted directly in bank or through subsidiaries. No firewalls mandated.	Permitted; bank may only directly sell insurance products/services. However, both banks and insurance firms are allowed to form "concern constellation" (financial groups) as long as the two activities are conducted in different firms.	Restricted; generally limited to bank premises.
United Kingdom Bank of England	Unrestricted; conducted either directly in bank or through subsidiaries. However, gilt-edged market making must be conducted through a subsidiary. No firewalls mandated.	Permitted; sales of insurance products/services may be conducted directly in bank, but underwriting only through subsidiaries. However, the bank's investment in the subsidiary must be deducted from the bank's capital when calculating its capital adequacy if the bank ownership share in the subsidiary exceeds 20%.	Unrestricted; conducted either directly in bank or through subsidiaries.

Figure 13. Permissible banking activities and bank ownership in selected EU and G-10 countries: 1995

Country and Bank Supervisor(s)	Securities[1]	Insurance[2]	Real Estate[3]
United States Federal Reserve System, Comptroller of the Currency, Federal Deposit Insurance Corporation, and State Authorities	Restricted; national and state member banks generally are prohibited from underwriting or dealing in corporate debt and equity instruments or securities. They may, however, engage in discount and full service brokerage as well as serve as agent for issues in privately placing securities. State non-member banks are subject to the same restriction as national banks, unless the FDIC determines the activity would not pose a significant risk to the deposit insurance fund. Bank holding companies may, on a case by case basis, be permitted to underwrite and deal in corporate debt and equity securities through a Section 20 subsidiary so long as the subsidiary's revenues for these activities do not exceed 10 percent of total gross revenues. Firewalls are mandated.	Restricted, banks generally may engage in credit life and disability insurance underwriting and agency activities. National banks, in addition, bay engage in general insurance agency activities in towns with less than 5,000 in population	Restricted; banks generally are restricted to investment in premises or that which is necessary for the transaction of their business.
European Union [4]	Not applicable; permissibility is subject to home country authorization and limited host country regulation, primarily notification requirements. (A single EU "passport" exists.)	Not applicable; permissibility is subject to home country regulation	Not applicable; permissibility is subject to home country and host country regulation.

Figure 13. Permissible banking activities and bank ownership in selected EU and G-10 countries: 1995

Country and Bank Supervisor(s)	**Commercial Bank Investment in Non-financial Firms**	**Non-financial Firm Investments in Commercial Banks**	**Geographical Branching Restrictions on Commercial Banks within Country**		
			Domestic Banks	*Non-Domestic Banks*	*Prior Regulatory Approval Required*
France Credit Institutions Committee, Bank Regulatory Commission, and Banking Commission	Unrestricted; Complies with EC Second Banking Directive. Subject to this limitation a bank may own 100% or the equity in any non-financial firm. [5]	Unrestricted, complies with the EC Second Banking Directive. [6]	None	None	No
Germany Federal Banking Supervisory Office and Deutsche Bundesbank	Unrestricted; Complies with EC Second Banking Directive. Subject to this limitation a bank may own 100% or the equity in any non-financial firm.[5]	Unrestricted, complies with the EC Second Banking Directive. [6]	None	None	No
Italy Bank of Italy	Restricted; more restrictive than the EC Second Banking Directive. Most banks are subject to an overall investment limit of 15% of own funds (7.5% in the case of unlisted firms) and to a concentration limit of 3% of own funds in each holding in non-financial firms or groups. Some banks, due to their size and proven stability, are subject to less stringent limits (overall and concentration limits of respectively 50% and 6% or leading banks, and 60% and 15% for specialized banks). Consistency with the principle of separation between banking and commerce is ensured by a further investment limit of 15% of invested firms' capital for all banks. [5]	Restricted; more restrictive than the EC Second Banking Directive. Persons who engage in significant business activity in sectors other than banking and finance are forbidden from acquiring an equity stake which, when added to those already held, would result in a holding exceeding 15% of the voting capital of a bank or in control of the bank. [6]	None	None	No
Japan Ministry of Finance (primary responsibility and Bank of Japan	Restricted; a single bank's ownership is limited to 5% of a single firm's shares, including other banks (Article 9, Anti-Monopoly Law).	Restricted; total investment is limited to firms capital or net assets. The Anti-Monopoly Law prohibits establishment of a holding company whose main business is to control the business activities of other domestic companies through the holding of ownership.	None	None	No

Figure 13. Permissible banking activities and bank ownership in selected EU and G-10 countries: 1995

Country and Bank Supervisor(s)	**Commercial Bank Investment in Non-financial Firms**	**Non-financial Firm Investments in Commercial Banks**	**Geographical Branching Restrictions on Commercial Banks within Country**		
			Domestic Banks	*Non-Domestic Banks*	*Prior Regulatory Approval Required*
Sweden Financial Supervisory Authority	Restricted; Investments on an aggregated basis are limited to 40% of a bank's own funds. Ownership in a firm is limited to 5% of this base (i.e. 1.5% in a firm or group of firms related to each other). Furthermore, ownership in a firm must not exceed 5% of the total voting power in the firm concerned. These limits do not apply when a bank has to protect itself against credit losses. In this case the bank must sell when market conditions are appropriate.[5]	Restricted; ownership is limited to 50% except under certain circumstances when a bank is near insolvency and there is a need for external capital injection. In the latter case, greater ownership may be permitted, based upon suitability of new owners. [6]	None	None	Yes
United Kingdom Bank of England	Unrestricted; complies with the EC Second Banking Directive. Subject to this limitation, a bank may own 100% of the equity in any non-financial firm. However, an ownership share of more than 20% requires that the investment be deducted from the bank's capital when calculating its capital adequacy on a risk basis. Otherwise, the investment is treated as a commercial loan for the risk-based calculation.	Unrestricted; complies with the EC Second Banking Directive. However, a firm would have to make application to the Bank of England to become a shareholder controller and receive the Bank's non-objection.	None. But need to comply with the local requirements and have adequate systems and controls for the function.	None. However, a bank must make an application to open a branch unless passporting into the UK under the EC Second Banking Directive.	Yes (see adjacent column).
United States Federal Reserve System, Comptroller of the Currency, Federal Deposit Insurance Corporation, and State Authorities	Restricted; national and state member banks generally are prohibited from making direct equity investments in voting or nonvoting stock. State non-member banks generally are limited to investments that are permissible for national banks. Bank holding companies are limited to an investment not to exceed 25% of a non-financial firm's capital.	Restricted; a non-financial firm may make equity investments in banks and bank holding companies. However, the investment must not exceed 25% of the bank's capital to avoid becoming a bank holding company. In other words, banks may only be acquired by companies that limit their activities to those deemed to be closely related to banking by the Federal Reserve Board	Yes	Yes; same restrictions that apply to domestic banks.	Yes.

Figure 13. Permissible banking activities and bank ownership in selected EU and G-10 countries: 1995

	EU Banks	**Non-EU Banks**	*Domestic Banks*	*Non-Domestic Banks*	*Prior Regulatory Approval Required*
European Union[4]	Unrestricted; the EC Second Banking directive (Article 12) limits "qualifying investments" to no more than 15% of a bank's own funds for investment in a single firm, and to no more than 60% for all investment in non-financial firms. In exceptional circumstances, these limits may be exceeded, but the amount by which the limits are exceeded must be covered by a bank's own funds and these own funds may not be included in the solvency ratio calculation. A qualifying investment is defined as a direct or indirect holding in an undertaking equal to at least 10% of its capital or voting rights or permitting the exercise of significant influence over its management.	Unrestricted; subjects qualifying investments to regulatory consent based only on the suitability of shareholders.	None (A single EU "passport" exists.)	Restricted; branches are fully regulated by the authorities of the EU member state in which they are situated and do not have access to the single EU "passport" to provide services or establish subsidiary branches throughout the EU.	

Source: Supervisory authorities in the listed countries provided information used to prepare this table. However, they are not responsible for any errors or misinterpretations. For exact information, one must consult the pertinent laws and regulations in the individual countries. For France and Japan, a source was Institute of International Bankers (1995).

Definitions**:** Unrestricted – A full range of activities in the given category can be conducted directly in the bank.
Permitted – A full range of activities can be conducted, but all or some must be conducted in subsidiaries.
Restricted – Less than a full range of activities can be conducted in the bank or subsidiaries
Prohibited – The activity cannot be conducted in either the bank or subsidiaries.

[1] Securities activities include underwriting, dealing and brokering all kinds of securities and all aspects of the mutual fund business.
[2] Insurance activities include underwriting and selling insurance products/services as principal and as agent.
[3] Real Estate activities include investment, development and management.
[4] The EU members are Austria (January 1, 1995), Belgium (original member), Denmark (January 1, 1973), Finland (January 1, 1995), France (original member), Germany (original member), Greece (January 1, 1981), Ireland (January 1, 1973), Italy (original member), Luxembourg (original member), Portugal (January 1, 1986), Spain (January 1, 1986), Sweden(January 1, 1995), and the United Kingdom (January 1, 1973).
[5] The EC Second Banking Directive (Article 12) limits "qualifying investments" to no more than 15% of a bank's own funds for investments in a single non-financial firm and to no more than 60% for aggregate investments in non-financial firms. In exceptional circumstances these limits may be exceeded, but the amount by which the limits are exceeded must be covered by a bank's own funds and these own funds may not be included in the solvency ratio calculation. A qualifying investment is defined as a direct or indirect holding in an undertaking equal to at least 10% of its capital or voting rights or permitting the exercise of significant influence over its management.
[6] The EC Second Banking directive (Article 11) subjects qualifying investments to regulatory consent based only on the suitability of shareholders.

Figure 14. Permissible banking activities and bank ownership in the EU and G-10 countries: 1995

	Securities	**Insurance**	**Real Estate**	**Commercial Bank Investment in Nonfinancial firms**	**Nonfinancial Firm Investment in Commercial Banks**
Very Wide Powers					
Austria	Unrestricted	Permitted	Unrestricted	Unrestricted	Unrestricted
Switzerland	Unrestricted	Permitted	Unrestricted	Unrestricted	Unrestricted
United Kingdom	Unrestricted	Permitted	Unrestricted	Unrestricted	Unrestricted
France	Unrestricted	Permitted	Permitted	Unrestricted	Unrestricted
Netherlands	Unrestricted	Permitted	Permitted	Unrestricted	Unrestricted
Wide Powers:					
Denmark	Unrestricted	Permitted	Permitted	Permitted	Unrestricted
Finland	Unrestricted	Restricted	Permitted	Unrestricted	Unrestricted
Germany	Unrestricted	Restricted	Permitted	Unrestricted	Unrestricted
Ireland	Unrestricted	Prohibited	Unrestricted	Unrestricted	Unrestricted
Luxembourg	Unrestricted	Permitted	Unrestricted	Unrestricted	Restricted
Portugal	Unrestricted	Permitted	Restricted	Permitted	Unrestricted
Spain	Unrestricted	Permitted	Restricted	Unrestricted	Permitted
Somewhat Restricted Powers:					
Italy	Unrestricted	Permitted	Restricted	Restricted	Restricted
Sweden	Unrestricted	Permitted	Restricted	Restricted	Restricted
Belgium	Permitted	Permitted	Restricted	Restricted	Unrestricted
Canada	Permitted	Permitted	Permitted	Restricted	Restricted
Greece	Permitted	Permitted	Restricted	Unrestricted	Unrestricted
Restricted Powers:					
Japan	Restricted	Prohibited	Restricted	Restricted	Restricted
United States	Restricted	Restricted	Restricted	Restricted	Restricted

Source: Figure 13.

Notes**:** Securities activities include underwriting, dealing and brokering all kinds of securities and all aspects of the mutual fund business.
Insurance activities include underwriting and selling insurance products/services as principal and as agent.
Real estate activities include investment, development and management.

Definitions**:** Unrestricted – A full range of activities in the given category can be conducted directly in the bank.
Permitted – a full range of activities can be conducted, but all or some must be conducted in subsidiaries.
Restricted – Less than a full range of activities can be conducted in the bank or subsidiaries.
Prohibited – The activity cannot be conducted in either the bank or subsidiaries.

Figure 15. Permissible corporate organizational form in which to conduct selected bank activities in selected EU and G-10 countries*

		Securities Activities[1]			
Country	**Bank Holding Company Permitted**	**Directly in the Bank**	**Bank Subsidiary**	**Bank Holding Company Subsidiary**	**Most Frequently Conducted in**
Austria	Yes, but infrequently used	Yes	Yes	Yes	Bank[4]
Canada	No	No	Yes	No	Bank Subsidiary
Finland	Yes, but infrequently used	Yes	Yes	Yes	Bank
Germany	Yes, but infrequently used	Yes	Yes	Yes	Bank
Greece	No[8]	Yes[9]	Yes	No	Bank Subsidiary
Ireland	Yes, but infrequently used	Yes	Yes	No	Bank Subsidiary
Italy	Yes, widely used	Yes	Yes	No	Bank
Luxembourg	No[14]	Yes	Yes	No	Bank
Netherlands	Yes, widely used	Yes	Yes	Yes	Bank
Portugal	Yes, but infrequently used	Yes	Yes	Yes	Bank & Bank Subsidiary
Spain	Yes, but infrequently used	Yes	Yes	NA	Bank Subsidiary & Bank Subsidiary[15]
Sweden	No	Yes	Yes	No	Bank
Switzerland	Yes, but infrequently used	Yes	Yes	Yes	Bank
United Kingdom	Yes, but infrequently used	Yes	Yes	Yes	Varies

Figure 15. Permissible corporate organizational form in which to conduct selected bank activities in selected EU and G-10 countries*

	Insurance Activities[2]			
Country	**Directly in the Bank**	**Bank Subsidiary**	**Bank Holding Company Subsidiary**	**Most Frequently Conducted in**
Austria	No	No	Yes	Bank Holding Company Subsidiary[5]
Canada	No	Yes	No	Bank Subsidiary
Finland	Yes & No [6]	Yes	Yes	Bank Subsidiary
Germany	No[7]	Yes	Yes	Bank Subsidiary
Greece	Yes[10]	Yes	No	Bank Subsidiary
Ireland	Yes[12]	Yes[12]	No	Bank
Italy	Yes	Yes	Yes	Bank Subsidiary[13]
Luxembourg	No	Yes	No	Bank Subsidiary
Netherlands	No	Yes	Yes	Bank Holding Company Subsidiary
Portugal	Yes	Yes	Yes	Bank & Bank Subsidiary
Spain	No	Yes	NA	Bank Subsidiary
Sweden	No	Yes	No	Bank Subsidiary
Switzerland	Yes	Yes	Yes	Bank Subsidiary
United Kingdom	Yes	Yes	Yes	Bank Subsidiary[16]

Figure 15. Permissible corporate organizational form in which to conduct selected bank activities in selected EU and G-10 countries*

	Real Estate Activities[3]			
Country	**Directly in the Bank**	**Bank Subsidiary**	**Bank Holding Company Subsidiary**	**Most Frequently Conducted in**
Austria	Yes	Yes	NA	Bank
Canada	Yes	Yes & No	No	Bank Subsidiary
Finland	No	Yes	No	Bank Subsidiary
Germany	Yes	Yes	Yes	Bank Subsidiary
Greece	No[11]	Yes	No	Bank Subsidiary
Ireland	Yes	Yes	No	Bank
Italy	No	Yes	Yes	Bank Subsidiary
Luxembourg	Yes	Yes	No	Bank Subsidiary
Netherlands	No	Yes	Yes	Bank Subsidiary and Bank Holding Company Subsidiary
Portugal	No	Yes	Yes	Bank Subsidiary
Spain	No	Yes	NA	Bank Subsidiary
Sweden	No	No	No	NA
Switzerland	Yes	Yes	Yes	Bank Subsidiary
United Kingdom	Yes	Yes	Yes	Varies

Source: Office of the Comptroller of the Currency using information provided by bank supervisory authorities in the respective countries.

* Information as of January 1997.

1 Securities activities include underwriting, dealing and brokering all kinds of securities and all aspects of the mutual fund business.

2 Insurance activities include underwriting and selling insurance products/services as principal and as agent.

3 Real estate activities include investment, development and management.

4 Securities activities fall under the banking activities provisions of Section 1 Austrian Banking Act. Hence, such business may be conducted exclusively by a bank.

5 Insurance activities require a license by the insurance supervisory authority (Ministry of Finance).

6 Insurance activities in Finland may be conducted in the bank as agent but not as principal.

7 Except as agent for insurance companies.

8 Holding companies may own the majority of shares in a Greek bank, but there is no specific legal framework referring to such companies.

9 Only underwriting and custodian services.

10 Only selling insurance products combined with deposits – no insurance risk may be assumed by banks.

11 Excluding investment in bank premises.

12 Only includes selling insurance products and services as agent.

13 Italian banks are not directly involved in insurance activities; these must be conducted by insurance companies subject to specific rules. Banks usually act as an agent of insurance companies, selling product through their branches.

[14] Pure holding companies are permitted to incorporate under Luxembourg law, but the statute of a bank holding company does not exist. This type of company is not subjected to any prudential control by any authority.

[15] Public debt directly in bank and stock exchange in bank subsidiary.

[16] With the exception of selling insurance as an agent, which is commonly conducted directly in the bank.

References

Allen, F. and D. Gale, (1988), ”Optimal Security Design”, *Review of Financial Studies*, Vol. 1(3), p. 229-263.

Allen, Franklin and Anthony M. Santomero, (1997), ”The Theory of Financial Intermediation”, *Journal of Banking & Finance*, Vol. 21 (11-12), p. 1461-85, December.

Bank for International Settlements, (1996), ”Settlement Risk in Foreign Exchange Transactions”, Basle, March (The Allsopp Report).

Bank for International Settlements (1998), ”Reducing Foreign Exchange Settlement Risk: A Progress Report”, Basle, July.

Barth, James R., Daniel E. Nolle, Tara N. Rice, (1997), ”Commercial Banking Structure, Regulation, and Performance: An International Comparison”, Office of the Comptroller of the Currency Working Paper, March.

Benston, George J. and George G. Kaufman, (1988), *Risk and Solvency Regulation of Depository Institutions: Past Policies and Current Options*, Monograph Series in Finance and Economics, New York University, Salomon Center for Graduate School of Business.

Benston, G.J., D. Brumbaugh, J. Guttentag, R. Herring, G. Kaufman, R. Litan, and Ken Scott, (1989), *Blueprint for Restructuring America’s Financial Institutions*, Washington D.C., The Brookings Institution.

Bernanke, B.S. and M. Gertler, (1989), ”Agency Costs, Net Worth and Business Fluctuations”, *American Economic Review*, Vol. 1, p. 14-30.

Bernanke, B.S. and M. Gertler, (1990), ”Financial Fragility and Economic Performance”, *The Quarterly Journal of Economics*, Vol. 3, p. 559-583.

Bhattacharya, S. and A. V. Thakor, ”Contemporary Banking Theory”, *Journal of Financial Intermediation*, Vol. 3, 1993, p. 2-50.

Black, Fischer, (1975), ”Bank Funds Management in an Efficient Market”, *Journal of Financial Economics*, 2, December, p. 323-39.

Boyd, John and Mark Gertlter, (1994), ”Are Banks Dead? Or are the Reports Greatly Exaggerated?”, Federal Reserve Bank of Minneapolis, *Quarterly Review*, Summer.

Caprio, Gerard, Jr., (1996), ”Bank Regulation: Case of the Missing Model”, Finance and Private Sector Development Division, The World Bank, Policy Research Working Paper No. 1574, January.

Davies, Howard, (1998), ”Why Regulate?”, Henry Thornton Lecture, City University Business School, November 4.

Diamond, D. W., (1984), ”Financial Intermediation and Delegated Monitoring”, *Review of Economic Studies*, Vol. 51, p. 393-414.

Diamond, D. W. and P. H. Dybvig, (1983), ”Bank Runs, Deposit Insurance, and Liquidity”, *Journal of Political Economy*, Vol. 1, p. 88-11.

Ellis, David M. and Flannery, Mark J., (1992), ”Does the Debt Market Assess Large Banks’ Risk?”, Time Series Evidence from Money Center CDs, *Journal of Monetary Economics*, Vol. 30 (3), p. 481-502, December.

European Commission, Communication of the Commission, ”Financial Services: Building a Framework for Action”, October 31, 1998.

Fama, Eugene F., (1985), ”What’s Different About Banks?”, *Journal of Monetary Economics*, 15, p. 29-39.

Flannery, Mark J., (1998), ”Modernizing Financial Regulation (Again)”, paper presented at a Conference on Financial Modernization and Regulation, Federal Reserve Bank of San Francisco, September 18.

Flannery, Mark, Sorescu, Sorin M., (1995), ”Evidence of Bank Market Discipline in Subordinated Debenture Yields: 1983-1991”, *Journal of Finance*, Vol. 50 (2), p. 699-718, June.

Furlong, Frederick T, Keeley, Michael C., (1987), ”Bank Capital Regulation and Asset Risk”, *Federal Reserve Bank of San Francisco Economic Review*, Vol. 0 (2), p. 20-40, Spring.

Furlong, Frederick T., Keeley, Michael C., (1991), ”Capital Regulation and Bank Risk-Taking: A Note”, *Federal Reserve Bank of San Francisco Economic Review*, Vol. 0 (3), p. 34-39, Summer.

Goodfriend, Marvin S., (1989), ”Money, Credit, Banking and Payment System Policy”, in *The U.S. Payments System: Efficiency Risk and the Role of the Federal Reserve*, edited by D. Humphrey, Kluwer Academic Press.

Gorton, G., (1988), ”Banking Panics and Business Cycles”, *Oxford Economic Papers*, Vol. 40, p. 751-781.

Gorton, G. and G. Pennacchi, (1990), ”Financial Intermediaries and Liquidity Creation”, *Journal of Finance*, Vol. 45, p. 49-71.

Gorton, Gary and A. Santomero, Anthony M., (1990), ”Market Discipline and Bank Subordinated Debt”, *Journal of Money, Credit & Banking*, Vol. 22, (1), p. 119-28, February.

Global Derivatives Study Group, (1993), *Derivatives: Practices and Principles*, Group of Thirty Report.

Guttentag, Jack M. and Richard J. Herring, (1988), ”Prudential Supervision to Manage Systemic Vulnerability”, *Financial Services Industry in the Year 2000: Risk and Efficiency*, Proceedings of the Conference on Bank Structure and Competition, Federal Reserve Bank of Chicago, p. 502-633.

Herring, Richard J. and Anthony M. Santomero, (1991), *The Role of Financial Structure in Economic Performance*, (in Swedish) SNS Forlag, Stockholm.

Humphrey, D. B., ”Payments Finality and Risk of Settlement Failure”, *Technology and the Regulation of Financial Markets*, edited by A. Saunders and L.J. White, Lexington: D.C. Heath, 1986, p.97-120.

Jacklin, Charles J., (1987), ”Demand Deposits, Trading Restrictions, and Risk Sharing”, ”Contractual Arrangements for Inter-temporal Trade”, Prescott, Edward C., ed. Wallace, Neil, ed., *Minnesota Studies in Macroeconomics Series*, Vol. 1 Minneapolis, University of Minnesota Press, p. 26-47.

Kane, Edward J., (1995), ”Three Paradigms for the Role of Capitalization Requirements in Insured Financial Institutions”, *Journal of Banking & Finance*, 19, p. 431-459.

Kareken, John H. and Neil Wallace, (1978), ”Deposit Insurance and Bank Regulation: A Partial-Equilibrium Exposition”, *Journal of Business*, Vol. 51 (3), p. 413-38, July.

Kuester, K.A. and J.M. O’Brien, (1990), ”Market based deposit insurance premiums”, Proceedings of the Conference on Bank Structure and Competition, Federal Reserve Bank of Chicago, p. 62-95.

Litan, R., (1991), *The Revolution in US Finance*, Brookings Institution.

Herring, Richard J. and Robert E. Litan, (1995), *Financial Regulation in the Global Economy*, Washington D.C., The Brookings Institution.

Horvitz, Paul, (1983), ”Deposit Insurance after Deregulation”, *Proceedings of the 9th Annual Conference of the Federal Home Loan Bank of San Francisco*, December.

Lewis, M. K., (1991), ”Theory and Practice of the Banking Firm”, in C. Green and D.T. Llewellyn, eds., *Survey of Monetary Economics*, Vol. 2, London, Blackwell Press.

Litan, Robert E., (1987), *What Should Banks Do?*, Washington D.C., The Brookings Institution.

Llewellyn, David, (1999), *The Economic Rationale for Financial Regulation*, Occasional Paper Series 1, London: Financial Services Authority, April.

Merton, Robert C., (1989), ”On the Application of the Continuous-Time Theory of Finance to Financial Intermediation and Insurance”, *The*

Geneva Papers on Risk and Insurance, Vol. 14, Number 52, July, p. 225-261.

Merton, Robert C., (1990), ”The Financial System and Economic Performance”, *Journal of Financial Services Research*, Volume 4, Number 4, (December), p. 263-300.

Miller, Merton H., (1995), ”Do the M&M propositions apply to banks?”, *Journal of Banking and Finance*, 19 (June), p. 483-489.

Pierce, James L., (1991), *The Future of Banking*, New Haven, Yale University Press.

Santomero, Anthony M., (1997a), ”Deposit Insurance: Do We Need It and Why?”, *Ekonomia*, Vol. 1 (1), p. 1-19, Summer.

Santomero, Anthony M., (1997b), ”Effective Financial Intermediation”, *Policy-Based Finance and Alternatives: East Asian Lessons for Latin America and the Caribbean*, Kim B. Staking, editor, Inter-American Development Bank.

Santomero, Anthony M., (1984), ”Modeling the Banking Firm”, *Journal of Money, Credit, and Banking*, 16, Part 2, p. 576-602,.

Santomero, Anthony M., ”The Bank Capital Issue”, *Financial Regulations and Monetary Arrangements after 1992*, M. Fratianni, C. Wihlborg, T.D. Willet eds., Economic Analysis Series, New Holland Press, 1991.

Santomero, Anthony M., ”The Regulatory and Public Policy Agenda for Effective Intermediation in Post Soviet Economies”, *in Financial Sector Reform and Privatization In Transition Economies* (7), p. 153-175, J. Doukas, V. Murinde and C. Wihlborg, editors, Elsevier Science Publishers B.V., 1998.

Santomero, Anthony M. and David Babbel, *Financial Markets, Instruments* and Institutions, Irwin Publications, 1996, Irwin Publications, 1997.

Santomero, Anthony M. and Paul Hoffman, (1998), ”Life Insurance Firms in the Retirement Market: Is the News All Bad?”, *Journal of the American Society of CLU and ChFC*, July.

Santomero, Anthony M. and Jeffrey J. Trester, (1998), ”Financial Innovation and Bank Risk Taking,” *Journal of Economic Behavior & Organization*, Vol. 35 (1), p. 25-37. March.

Santomero, Anthony M. and Jeffrey J. Trester, (1997), ”Structuring Deposit Insurance for a United Europe,” *European Financial Management Journal*, June.

Consolidation and Strategic Positioning in Banking with Implications for Sweden

by

Arnoud W. A. Boot*

Supplement number 27 to the Government Inquiry on the International Competitiveness of the Swedish Financial Sector.

Views, suggestions, and opinions expressed in this study represent the personal views of the author and not necessarily those of the Government Inquiry on the International Competitiveness of the Swedish Financial Sector.

May, 1999

* University of Amsterdam, Faculty of Economics and Econometrics, Finance Group, Roetersstraat 11, 1018 WB, Amsterdam, The Netherlands, E-mail: awaboot@fee.uva.nl. The author thanks Jonas Niemeyer, Pehr Wissén and participants at the Symposium on the International Competitiveness of the Swedish Financial Services Industry (Stockholm, March 1999).

Table of Contents

Kort sammanfattning (in Swedish)

I bilagan görs en genomgång av de faktorer som anses betydelsefulla när det gäller stordriftsfördelar etc. i bankverksamhet och hur detta påverkar strukturomvandlingen. Olika empiriska undersökningar kring detta refereras också. Avslutningsvis förs mot den bakgrunden en diskussion om utvecklingen i den svenska banksektorn.

Bilagan konstaterar att det för närvarande pågår en snabb strukturomvandling på bankområdet i nästan alla länder, däribland Sverige, med större banker och ökad koncentration som ett viktigt inslag i bilden. Författaren menar dock att denna process bara i begränsad utsträckning kan förklaras med viljan att ta tillvara skalfördelar eller av andra produktionsekonomiska drivkrafter. Den dominerande drivkraften förefaller istället vara en önskan att positionera sig strategiskt i den snabba omvandling som pågår och där osäkerheten om den framtida utvecklingsriktningen och de framtida utvecklingsförutsättningarna är stor. En konsolidering kan här ge ett större manöverutrymme inför kommande faser av strukturförändringar.

1 Introduction

The financial services industry is restructuring and consolidating at an unprecedented pace around the globe. Particularly, in the United States and Western Europe transactions are numerous and breathtaking. But restructuring is also going on in Asia. Most striking is probably the ever-escalating scale of mergers in banking. In just the last few years, in the US mergers have led to a consolidation of money center banks (e.g. the Chase Manhattan and Chemical Bank merger) and the emergence of regional power houses (e.g., the expansion strategies of BankOne and Nationsbank, and their subsequent mergers with, respectively, First Chicago/NBD and BankAmerica). In Europe, mergers have also been prominent. While cross border mergers are relatively infrequent – with exceptions in Scandinavia and the acquisition of the Belgian Bank BBL by the Dutch financial conglomerate ING – mergers between domestic institutions typically involve large universal banks and are often spectacular. Noteworthy examples include the marriage of the Union Bank of Switzerland and Swiss Bank Corporation and the announced merger between Société Général and Paribas (and possibly BNP). And in Japan, a spectacular merger has produced the new Tokyo-Mitsubishi bank with over $ 700 billion in assets.

A parallel phenomenon is the broadening of scope of many banks. Even banks that traditionally followed well motivated focused strategies now seem to give in to this trend. For example, Bankers Trust with its activities aimed at the corporate market, now puts itself in the arms of a scope expanding universal bank (Deutsche Bank). Scope-expansion also originates from investment banks. Major investment banks are redefining their domain by offering traditional commercial banking products like commercial and industrial loans and by moving into retail brokerage. The union of Salomon Brothers (investment bank) and Smith Barney (brokerage) within Travelers underscores the scope-expansion in the industry. The spectacular cross-industry merger by Citicorp and Travelers also brings the insurance activities together with bank-oriented financial services. This concept is not really new. Some European banks, – e.g. ING in the Netherlands – already engage in *bancassurance*, that is, combining banking and insurance activities. Similarly, Credit Suisse expanded into insurance by acquiring the insurance corporation Winterthur.

One question is then immediate. Why are banks consolidating so much and expanding scope? The popular financial press points to the increasingly competitive environment of banking as the culprit. As commercial banking becomes more competitive, banks need to examine all possible ways to wring inefficiencies out of their cost structures. One way to do this is to merge with other banks and realize efficiencies of scale through elimination of redundant branches and back-office consolidation. Moreover, the diminishing margins in commercial banking invite banks to look outside their traditional domain. Some non-banking activities may offer higher margins and make scope expansion attractive. These higher margins may come in part from the value customers attach to "one-stop shopping".

However, these popular explanations are inadequate. The empirical evidence on scale and scope economies in banking is far from conclusive.[1] It is questionable whether these economies are large enough to justify banking consolidation and scope expansion (see Berger [1997] and Berger, et al. [1993]). Moreover, ample research in corporate finance points at the existence of a "diversification discount". On average diversification does seem to destroy value. There is also evidence that improvements in operating performance and stock returns have been experienced by firms that have refocused (see John and Ofek [1995] and Comment and Jarrell [1995]). Therefore, the important question is why are there so many mergers and acquisitions taking place in the industry?

This study aims to address this question and other related issues. I will examine the existing empirical evidence on scope and scale economies in banking. A recent survey paper by Berger, Demsetz and Strahan [1999] is of substantial help. An important question is whether the existing empirical evidence can be used to explain the current consolidation wave. While I will conclude that the existing evidence is of some value, I doubt that it is really helpful for understanding the current restructuring in banking. Several issues play a role here. Apart from econometric and sample-selection issues, and possibly fundamental changes in underlying "state-variables", *the* important issue is that *strategic* considerations are the driving force behind the current consolidation wave. As I will argue, these considerations may have little to do with true scale or scope economies. Rather learning, first-mover advantages and strategic advantages of market power and associated "deep pockets" may explain the current consolidation wave. Strategic positioning might be the rule of the game, and be an optimal response to

[1] See Shaffer and David [1991], Cornett and Tehranian [1992], Mester [1992], Mitchell and Onvural [1996] and Clark [1996].

the uncertainties and rapid (and unpredictable) changes facing financial institutions today. Consolidation might then be an evolutionary phenomenon and be followed by a new type of repositioning when the uncertainties become more manageable. This analysis follows recent co-authored work with Todd Milbourn and Anjan Thakor (Boot, Milbourn and Thakor [1999]).

The organization of this study is as follows. In Section 2, I start out with a discussion of the growing research in the field of financial intermediation. This research - mainly theoretical in nature – sheds light on the costs and benefits of bank funding vis-à-vis direct funding in the financial market. While primarily focused on the funding role of banks and financial markets, it provides valuable insights into the economics of banking. These insights are of great importance for understanding the role of financial institutions in the future. Subsequently, Section 3 discusses the extensive empirical literature on scale and scope economies in banking. Here, I will particularly look at scale and scope considerations that may be important in the future. An important issue in this context is that the literature needs to differentiate more between the various activities (services and products) of financial intermediaries. Section 4 introduces strategic considerations, in particular, the importance of *strategic positioning*; see the discussion above. Section 5 discusses some political considerations that may impact the restructuring of the financial services industry. In Section 6 I offer some thoughts on the future path of the ongoing restructuring. Foremost, I then address the options and prospects of the Swedish financial services industry. My message is an optimistic one, but one should be prepared for an ongoing extensive restructuring of the Nordic financial services industry.

2 Fundamentals: The Economics of Banking

2.1 Traditional versus modern banking

Traditional commercial banks hold non-marketable or illiquid assets that are funded largely with deposits. There is typically little uncertainty about the value of these deposits which are often withdrawable on demand. The liquidity of bank liabilities stands in sharp contrast to that of their assets, reflecting the banks' *raison d'être*. By liquifying claims, banks facilitate the funding of projects that might otherwise be infeasible.

The banks' assets are illiquid largely because of their information sensitivity. In originating and pricing loans, banks develop proprietary information. Subsequent monitoring of borrowers yields additional private information. The proprietary information inhibits the marketability of these loans. The access to information is the key to understanding the comparative advantage of banks. In many of their activities banks exploit their information and the related network of contacts. This relationship-oriented banking is a characteristic of value-enhancing financial intermediation. The relationship and network orientation does not only apply to traditional commercial lending but also to many areas of modern banking.

One might be tempted to interpret modern banking as transaction-oriented. So does an investment bank (IB) – generally considered a prime example of modern banking – facilitate a firm's access to public capital markets. The IB's role could be interpreted as that of a broker; i.e. matching buyers and sellers for the firms' securities. In this interpretation IBs just facilitate transactions, which would confirm the transaction orientation of modern banking. The IBs' added value would then be confined to their networks, i.e. their ability to economize on search or matching costs. As a characterization of modern banking, this would describe their economic role too narrowly. IBs do more. They – almost without exception – *underwrite* those public issues, i.e. absorb credit and/or placement risk. This brings an IB's role much closer to that of a commercial bank engaged in lending; the processing and absorption of

risk is a typical intermediation function similar to that encountered in traditional bank lending.

In lending, a bank manages and absorbs risk (e.g. credit and liquidity risks) by issuing claims on its total assets with different characteristics then those encountered in its loan portfolio. In financial intermediation theory this is referred to as *qualitative asset transformation* (see Greenbaum and Thakor [1995]). Underwriting of an IB can be interpreted analogically; risk is (temporarily) absorbed and is channeled through to the claim holders of the IB. The role of IBs is therefore more than just purely brokerage. Underwriting requires information acquisition about the borrower which is supported by a relationship orientation. A relationship orientation will therefore still be present in investment banking, both in the direction of investors ("placement capacity") and towards borrowing firms. Thus, characterizing financial market funding as transaction-oriented and bank lending as relationship-oriented is too extreme. What will be true, however, is that in investment banking relationships depend much less on local presence

2.2 Are bank loans special?

Some see public capital market financing as a potentially superior substitute for bank lending. This, however, is unwarranted. Bank lending has distinctive comparative advantages. In particular, it may support enduring close relationships between debtor and financier that may mitigate information asymmetries. This has several components. A borrower might be prepared to reveal proprietary information to its bank, while it would have never disseminated this information to the financial markets (Bhattacharya and Chiesa [1995]). A bank might also be more receptive to information because of its role as enduring and dominant lender. This amounts to observing that a bank might have better incentives to invest in information acquisition. While costly, the substantial stake that it has in the funding of the borrower, and its, hopefully, enduring relationship – with the possibility of information reusability over time – increase the value of information.[2]

The bank-borrower relationship is also less rigid than those normally encountered in the financial market. The general observation is that a better information flow facilitates more informative decisions. In

[2] Diamond [1984] introduces intermediaries as delegated monitors. See Chan, Greenbaum and Thakor [1986] for a discussion on information reusability, and James [1987] and Lummer and McConnell [1989] for empirical evidence. For a nice illustration supporting the special role of banks, see Berlin [1996].

particular, relationship finance could allow for more flexibility and possibly value-enhancing discretion. This is in line with the important ongoing discussion in economic theory on rules versus discretion, where discretion allows for decision making based on more subtle – potentially non-contractible – information.[3] Two dimensions can be identified. One dimension is related to the nature of the bank-borrower relationship. In many ways, it is a mutual commitment based on trust and respect. This allows for *implicit* – non-enforceable – long-term contracting. An optimal information flow is crucial for sustaining these "contracts". Information asymmetries in the financial market and the non-contractibility of various pieces of information may rule out long-term access to alternative capital market funding sources as well as *explicit* long-term commitments by banks. Therefore, both bank and borrower may realize the added value of their relationship, and have an incentive to foster their relationship.[4]

The other dimension is related to the structure of the explicit contracts that banks can write. Bank loans are generally easier to renegotiate than bond issues or other public capital market funding vehicles. The re-negotiation allows for a qualitative use of flexibility. Sometimes this is a mixed blessing because banks may suffer from a soft-budget constraint (the borrowers may realize that they can renegotiate ex post, which could give them perverse ex ante incentives). In reality, bank loans therefore often have *priority*. With priority a bank may strengthen its bargaining position and thus become tougher.[5] The bank could then credibly intervene in the decision process of the borrower when it believes that its long-term interests are in danger. For example, the bank might believe that the firm's strategy is flawed, or a restructuring is long overdue. Could the bank push for the restructuring? If the bank has no priority, the borrower may choose to ignore the bank's wishes. This is because the borrower realizes that the bank cannot enforce its demands. The bank could threaten to call the loan, but the borrower – anticipating the dreadful consequences not only for himself but also for the bank – realizes that the bank would never carry out such a threat. However, when the bank has priority, the prioritized claim may

[3] See e.g. Simon [1936] and Boot, Greenbaum and Thakor [1993].

[4] Mayer [1988] and Hellwig [1991] discuss the commitment nature of bank funding. Boot, Thakor and Udell [1991] address the *credibility* of commitments. Schmeits [1997] formally considers the impact of discretion (flexibility) in bank loan contracts on investment efficiency.

[5] See Dewatripont and Maskin [1995] on the issue of soft-budget constraints. Diamond [1993], Berglöf and Von Thadden [1993], and Gorton and Kahn [1993] address the priority structure.

insulate the bank from these dreadful consequences. It could now *credibly* threaten to call the loan, and enforce its wishes upon the borrower. This then identifies an important advantage of bank financing: *timely intervention*.[6]

These observations highlight the complementarity of bank lending and capital market funding. Prioritized bank debt facilitates timely intervention. This feature of bank lending is valuable to the firm's bondholders as well. They might find it optimal to grant bank debt priority over their own claims, and in doing so delegate the timely intervention activity to the bank.[7] Consequently, the borrower may reduce its total funding cost by accessing both the bank-credit market and the financial market.

The overall conclusion is that bank lending potentially facilitates more informative decisions based on a better exchange of information. While not universally valuable, this suggests a benefit of relationship banking.[8]

2.3 Securitization: A threat to bank lending?

Securitization is an example of a financial innovation – or an innovation in funding technology – that suggests a potential gain of (transaction-oriented) markets at the expense of bank lending. Is this true? Let's first evaluate the economics of securitization.[9]

Securitization is a process whereby assets are removed from a bank's balance sheet. Asset-backed securities rather than deposits would then fund dedicated pools of bank-originated assets. Securitization is an

[6] One could ask whether bond holders could be given priority and allocated the task of timely intervention. Note that bond holders are subject to more severe information asymmetries and are generally more dispersed (i.e. have smaller stakes). Both characteristics make them ill-suited for an 'early intervention task'.

[7] The bond holders will obviously ask to be compensated for their subordinated status. This – ignoring the timely intervention effect – is "a wash". In other words, the priority (seniority) or subordination features can be priced out. That is, as much as senior debt may *appear* cheaper (it is less risky), junior, or subordinated debt, will appear more expensive.

[8] See e.g. Petersen and Rajan [1994] and Houston and James [1995] for empirical evidence.

[9] Gorton and Pennachi [1995] provide an economic rationale for bank loan sales and securitization. See also Boot and Greenbaum [1995].

example of unbundling of financial services. More specifically, banks would no longer fund those assets, instead the investors buying the asset-backed securities would provide funding. As I will emphasize, securitization does not signal the demise of banks, even if it becomes an economically important innovation (and thus substantially reduces the banks' on-balance sheet assets). To see this point, one needs to analyze the traditional lending function in some detail.

The lending function can be decomposed into four more primal activities: origination, funding, servicing and risk processing. Origination subsumes screening prospective borrowers and designing and pricing financial contracts. Funding relates to the provision of financial resources. Servicing involves the collection and remission of payments as well as the monitoring of credits. Risk processing alludes to hedging, diversification and absorption of credit, interest rate, liquidity and exchange-rate risk. Securitization decomposes the funding activity; banks would no longer fund securitized assets.

The economics of securitization dictates that the originating bank *credit enhances* the issue. Credit enhancement is typically achieved through the provision of excess collateral or with a letter of credit. Effectively this means that the originating bank continues to bear part of the consequences (losses) if the securitized assets do not perform. The credit enhancement reduces the riskiness of the asset-backed claims from the investors' perspective, but, more importantly, it addresses conflicts of interest rooted in the originating bank's proprietary information. With private information in possession of the originating bank, the market requires assurances that the bank will truthfully reveal the quality of the assets it seeks to sell. As with a warranty in product markets, credit enhancement discourages misrepresentation by requiring the originator to absorb a portion of the losses owing to default. Similarly, credit enhancement signals the market that the originator will perform a thorough credit evaluation and an undiminished monitoring effort. Credit enhancement therefore reduces the information sensitivity of securitized claims by enhancing their marketability.[10]

Securitization could lead to a *reconfiguration* of banking. But even with widespread securitization the incremental value of banks would largely be preserved.[11] They would originate and service assets, while

[10] The reputation of the originating bank will be equally important. Moreover, accreditation by credit rating agencies could also add to the marketability of the securitized claims.

[11] See also Boyd and Gertler [1994]. They argue that banks have not lost importance. They argue that a substitution from on-balance sheet to off-balance sheet banking may have (falsely) suggested a shrinking role for

also processing the attendant risk in order to sustain these activities. Banks would therefore continue to screen and monitor borrowers, design and price financial claims, and provide risk management services.

How important will securitization become? I can only give a very tentative answer. So far, securitization barely exists in Europe. In the US securitization has spread rapidly in the last decade but mainly for car loans, mortgages and credit-card receivables. The standardization and modest size of these credits allows diversification of idiosyncratic risks upon pooling. Private information distortions – as discussed above in the context of credit enhancement – are thought to be less severe for these standardized credits. What does this imply for the larger, more customized and heterogeneous commercial loans? These tend to be more information sensitive. Their quality is therefore more dependent on the rigor of initial screening and subsequent monitoring. Hence, the pooling of commercial loans does less to dissipate their information sensitivity, attenuating the benefits of securitization.

These considerations, however, do not preclude the securitization of business credits. They merely elevate the cost. For example, with more information-sensitive assets, the originating bank may need to retain a larger portion of the credit risk; credit enhancement becomes more important. If the information sensitivity is too severe, credit enhancement, short of total recourse may not overcome the private-information problem. Thus, the potential advantages of securitization would largely be lost, and traditional bank lending would continue to dominate. However, for an increasing array of moderately information-sensitive assets, securitization might become the preferred intermediation technology. As my discussion of the economics of securitization suggests, banks even then continue to be indispensable for most of the primal activities that were previously combined together in bank lending. More importantly, the comparative advantage of banks rooted in proprietary information about their clientele would be preserved.

2.4 Is relationship banking at risk?

Relationships may facilitate a continuous flow of information between debtor and creditor which could guarantee a smooth access to funding. Banks' comparative advantages are often rooted in these relationships. Many believe that a competitive environment may threaten relationships. Borrowers might be tempted to switch to other banks, or to the financial

---

banks. As in the description of securitization in the text, much of the bank's value added in the primal activities would be preserved.

market. In particular, increased credit market competition imposes constraints on the ability of borrowers and lenders to inter-temporally share surpluses (Petersen and Rajan [1995]). When parties anticipate a shorter expected life-span of their relationships they may respond by reducing their relationship-specific investments. More specifically, anticipated shorter relationships inhibit the reusability of information, and thus diminish the value of information. Banks may then find it less worthwhile to acquire (costly) proprietary information, and relationships suffer. Paradoxically, shorter or weaker relationships actually become a self-fulfilling prophecy.

These arguments highlight the negative spiral that may undermine relationship banking. An important observation is that this negative spiral might be self-inflicted. While competitive banking challenges relationships, the bankers' response - cutting back on information acquisition - may actually damage relationship banking most.

Borrowers, however, face an equal challenge: how to benefit from competitive pricing without jeopardizing the benefits of relationships (see Rajan [1992])? This is the relationship puzzle. The relationship puzzle has no obvious solution. Relationships may foster the exchange of information, but may simultaneously give lenders an information monopoly and undermine competitive pricing.[12] Transaction-oriented finance, however, may give little incentive to acquire information but is potentially subjected to more competition. There might be no winners in this process; e.g. transaction-oriented finance may not be feasible where relationship-oriented finance retreats. More specifically, markets for transaction-oriented finance may fail when problems of asymmetric information are insurmountable. This argument is used by some to highlight the virtues of (relationship-oriented) bank-dominated systems (e.g. Germany and Japan) vis-à-vis market-oriented systems.[13]

As discussed in the preceding subsections, bank lending, securitization of loans and underwriting of public capital market issues may all benefit from a relationship orientation. The distinction between relationship-oriented finance and transaction-oriented finance, or between bank-dominated systems and market-oriented systems, may

[12] The informational monopoly on the ”inside” lender’s side may be smaller if a borrower engages in multiple banking relationships. This would mitigate the possibilities of rent extraction by informed lenders and induce more competitive pricing (see Sharpe [1990] and also Petersen and Rajan [1995].

[13] A fascinating academic literature is emerging on the design of financial systems. See Allen [1993], Allen and Gale [1995] and Boot and Thakor [1997]. One objective of this literature is to evaluate the pros and cons of bank-dominated and financial market-dominated systems.

therefore be less well-defined than it appears. What might be true is that a bank-dominated system invites oligopolistic behavior such that competition is contained (and relationships preserved) while a market-dominated system suppresses competition less.

A less competitive financial system may thus preserve relationships more. Competition threatens relationships, but it may simultaneously elevate the importance of relationships as a distinct competitive edge. This is the *relationship paradox*. A relationship orientation can alleviate competitive pressures. Thus, a more competitive environment should encourage banks to become client-driven, and customize services. Since a relationship orientation may earn banks a substantial added-value, banks could then isolate themselves from pure price competition.[14]

[14] Boot and Thakor [1999] develop this intuition further. They show that competition may indeed induce banks to divert resources to relationship-specific activities. In their model banks choose between 'passive' transaction lending and more intensive relationship lending. Transaction lending competes head-on with funding in the financial market. Their key result is that as interbank competition increases, banks make *more* relationship loans, but each has lower value-added for borrowers, relative to transaction loans. Capital market competition *reduces* relationship lending (and bank lending shrinks), but each relationship loan has greater value-added for borrowers. In both cases, welfare increases for some borrowers but not necessarily for all.

3 Scale and Scope Issues in Banking

Scale and scope economies are often cited as one of the main reasons behind the current merger and acquisition wave in banking. Are scale and scope economies present? And could they rationalize the current restructuring in the industry? Scale and scope economies in banking have been studied extensively. In general, the empirical evidence cannot readily identify substantial economies of scale or scope. Scale economies could not be found beyond a relatively small size of banks as measured by total assets (i.e., beyond $ 100 million up to $ 10 billion in total assets), see Table 1. Only recently have some studies succeeded in showing some economies of scale at a level of total assets up to $ 25 billion. Similar results were obtained for scope economies. These results seem hard to reconcile with the perceived wisdom of bankers and the observed mega-mergers.

Most empirical researchers in the area of industrial organization will acknowledge that scale and scope economies are difficult to measure. So, at best, very modest conclusions could ever be drawn from these empirical studies. With this in mind, we can start analyzing the evidence. A first observation is that the inconclusive results are not really surprising. Inefficiencies in managing larger organizations may mitigate possible scale and scope benefits. This would be in line with the sizable literature on the "diversification discount". A complication is also that increasing scale and scope may facilitate market power and thus elevate profitability in *absence* of scale and scope economies. This effect might be less important in inter-(geographic) market mergers. Moreover, alternative distribution network (e.g., direct banking) and the proliferation of financial markets may have reduced the effective market power of locally concentrated financial institutions. This points at a more general issue: the level of concentration may no longer be a good proxy for the (non-) competitiveness of a market. What more can be said?

Table 1. The empirical evidence on scale and scope economies

Subject	Main findings
• Market power analysis: effect on prices and profits	
→ Static	° Higher rates on small business loans, lower deposit rates ° Effect on deposit rates has become less in the 90's ° Deposit rates sticky (show downward rigidity) ° Multi-state BHC's charge higher fees to retail customers ° Small effect of concentration on bank profits ° Most profitable banks are *not* in most concentrated markets
→ Dynamic (effect of M&A)	° Downward effect on deposit rates if related to increase in concentration ° Effect on profitability ratio's mixed, but possibly positive ° Event studies show mixed results on combined value of target and acquirer: but focus (both geographic and activity) adds value. International mergers more profitable vis-à-vis domestic US mergers
• Efficiency consequences	
→ Static	° Main focus on cost-control: scale economies disappear between $ 100 million and $ 10 billion in total assets ° Scope economies studies find similar results to scale economies studies ° Recent 1990's studies find greater potential of scale economies (up to $ 25 billion in total assets, but possibly beyond) ° Potential for scale, scope and product mix efficiencies in managing risk, particularly diversification benefits of geographical expansion ° Combining bank and non-bank (insurance) activities has mixed effect on total risk
→ Dynamic	° Cost efficiency (controlled for market power effects) shows little improvement ° Efficiency gains in US mega-mergers, some for large in-market mergers ° Some diversification benefits: higher proportion of loans per dollar of capital (at expense of securities holdings)

Source: Berger, Demsetz and Strahan [1999].

Several other qualifications should be made. First, most studies concern the US Contrary to banking in many other countries, US banking has historically been quite fragmented.[15] The mergers and acquisitions that were included in most studies (mostly dating back to the 70's and 80's) took place in an environment where severe constraints existed on the type and geographic dispersion of activities. It is conceivable that these restrictions made it difficult to benefit from scale and scope economies (see also Calomiris and Karceski [1997]).

Second, the level of aggregation in most studies may obscure benefits to scale and scope. In particular, we should look at what type of mergers and acquisitions involve scale and/or scope benefits. For example, Flannery [1999] points at recent research that suggests that mergers with both a geographic and activity focus are most value enhancing.[16] Similarly, in analyzing scope and scale issues we should focus on the type of activities. What are the scale economies in each activity? And what product-mix offers true scope economies?[17] Some of these and other concerns are summarized in Table 2.

These observations offer yet little proof for true scale and scope economies. I see the following five primary sources of scale and scope economies (see also Canals [1994]):[18]

[15] This is not really surprising. U.S. banks face(d) substantial regulatory constraints on their activities concerning both the type of their activities (e.g. banks could engage in commercial banking or investment banking, not both) and their location (e.g. limits on interstate banking). More recently, however, regulatory constraints have become less binding. This undoubtedly partially explains the surge in mergers and acquisitions.

[16] An important issue is whether this only points at market-power benefits (see also Table 1) or whether also true efficiency gains could be at work.

[17] Surprisingly, this type of research is yet hard to find. A lot of research has been done on potential conflicts of interest in universal banking. To some extent, this is activity specific (investment banking versus commercial banking). However, this research is of very limited interest because it ignores the question of complementarity between activities. This is not really surprising because the literature is solely motivated by the obscure Glass-Steagall regulation in the U.S. (see Kroszner and Rajan [1994] and Puri [1996]).

[18] Observe that some of these sources of scope and scale economies are inter-related.

i. Information technology related economies;
ii. Distribution-network related benefits (strengthened by IT developments);
iii. Marketing/brand name and reputation related benefits;
iv. Financial-innovation related economies;
v. Benefits of diversification

Table 2. Some problems with the existing empirical studies on scale and scope economies

Subject	Issues
• Market power analysis: effect on prices and profits	
→ Static	° Is concentration the right measure? What about contestability of markets?
→ Dynamic (effect of M&A)	° Combined effect of market power and efficiency changes difficult to disentangle ° Profitability ratio's affected by market power ° Cost ratio's via costs of deposits linked to market power. Operational costs affected by relative importance of deposits versus purchased funds ° Event studies affected by "signaling". That is, the immediate effect of a merger announcement on stock prices incorporates all types of changes in expectations
• Efficiency consequences	
→ Static	° How to measure scope economies ° Lack of data points for mega-institutions.
→ Dynamic	° Little differentiation between type of merger and/or type of activities

The first source, information technology, is potentially of great importance. Most of the existing studies on scale and scope economies involve data that precede the information technology revolution. It is likely that recent information technology developments facilitate a much more efficient and effective utilization of information over ranges of services and customers. That is, client-specific information may allow for scope economies and facilitate a competitive advantage to financial institutions that can offer a range of services to their clientele. Similarly, possibilities for reusability of information across customers may have increased. Also, information technological developments may help

facilitate differentiation of products and services.[19] Together with the sizable investments projected in information technology, scale and scope benefits may have become more important. The implication is also that sizable investments in information technology are needed to truly benefit from scale and scope economies. This relates also to the second source: distribution network related benefits may be rooted in information technology developments. In particular, IT developments may facilitate scale economies offering gains in running a sizable distribution network.

The next source of scale and scope economies is linked to marketing and reputation. Marketing expenses involve substantial fixed costs. This suggests some economies of scale. Also, scope benefits may be present in the joint marketing of products to customers. Brand image is partially marketing related but is also linked to the notions of "trust", "reputation" and "confidence". These notions play an important role in the financial services industry. Increasingly, financial service providers offer services that crucially depend on their reputation. For example, the growing importance of off-balance sheet claims puts great emphasis on the ability of financial institutions to honor these *contingent* liabilities. Also, under certain conditions, increasing scale and scope allows financial institutions to capitalize more on their reputation. That is, a wider scope (or scale) may help a financial institution "to put its reputational capital at work" (see Boot, Greenbaum and Thakor [1993]).

The next source of potential scale and scope economies is financial innovation. Financial innovation as a source of scope and scale economies is a two-edged sword. Some suggest that larger institutions are less likely to innovate due to the inherent bureaucracy. This might be true but this is a governance issue. Ceteris paribus, larger institutions could better recoup the fixed costs of financial innovations. Innovations could be marketed to a larger customer base and/or introduced in a wider set of activities. For financial innovations scale and scope might be particularly important given the rapid imitation by competitors. Only for a short period of time does a true competitive advantage exist. A wider scope and larger scale may help recoup the fixed costs in this short period of time.

The last potential source of scale and scope economies is the benefit of diversification. From a corporate finance perspective, this benefit is controversial. After all, investors (shareholders) could diversify and why

[19] In Section 2.4, I have discussed the effect of increasing competition on the reusability of information. The conclusion there was that bank-borrower relations may have shortened in duration and hence reduced the reusability of information, but that simultaneously investing in information acquisition may have gained in importance for competitive reasons.

would a financial institution itself need to do this? But, nevertheless, low variability of returns is considered very important in banking.

My assessment is that scale and scope economies are present. However, the complexity of running the larger organizations needed to exploit these are far from trivial. I would expect, therefore, that the empirical evidence in a cross-section of financial institutions will continue to be mixed. In terms of observed bank strategies this will translate into the co-existence of specialized and more universal financial institutions.

4 Scope as a Strategic Advantage

4.1 General framework

The explanation developed in this section is that strategic uncertainty about future exploitable core competencies may dictate broadening of scope. The basic idea is as follows. Suppose a bank knows that – perhaps due to deregulation – it can participate in a non-banking market at some time in the future. The problem is that this is a new market for the bank, so the bank is highly uncertain about whether it has the skills to compete effectively in that market. It has two choices. It can wait until that future time to find out whether it has the capabilities and "core competencies" (as defined by Hamel and Prahalad [1990]) for this new market. Or it can enter the market "early" and discover what its skills are prior to making costly resource allocation decisions. The advantage of the second approach is that it permits the bank to "experiment" with a new business and learn whether it has the skills to compete in that business. This learning permits better decisions when competition commences. In particular, having better knowledge about its own skills allows the bank to be more aggressive in its output decisions and gain market share when it knows that its skills are superior to those of its competitors, and to exit the market when its skills are inferior.

One can explain scope expansion as the bank reserving the right to play in a variety of "new" activities. By making incremental investment today, the bank puts itself in a privileged position through the acquisition of superior information by learning. This allows the bank to wait until the environment becomes less uncertain before determining whether to compete in the new market and if so, how aggressively; see also Courtney et al. [1997] for the link between strategy and uncertainty.

In a recent paper (see Boot, Milbourn and Thakor [1999]) a formal model of banking has been developed that formalizes these ideas and incorporates scope as a potential competitive advantage. The framework in that paper is as follows. It starts out with a banking sector with narrowly defined existing activities and asks whether banks should expand into a "new" activity. A key feature of the analysis is that there is strategic future uncertainty about the demand for this new activity, i.e. the activity has prospects only in the long run and demand may not

materialize. The bank must decide whether or not to expand in this activity, and if so, whether to enter early or late. Early entry is costly because the activity becomes important only later. Demand may not materialize, and entering early requires investments to be made prior to the resolution of demand uncertainty. Moreover, the scope expansion associated with investing in strategic options could reduce the competitiveness of existing operations (say due to dilution of focus). However, early entry offers potential strategic advantages. In particular, early entry could lead to the discovery of skills that would allow for a more efficient delivery of the new activity and hence make the commercial bank a more credible competitor once the prospects of this activity become clear.

The question is: when will the benefits of early entry outweigh the costs? The uncertainty about skills plays a key role here. If this uncertainty is substantial, early entry may be beneficial. The other key factor is the competitive environment of the banking sector, and the anticipated competition for the new activity. Suppose that the new activity can also be offered by a specialized provider (a "boutique" specializing in this activity). If the commercial bank enters (early or late), we could consider the market for this activity as a Cournot duopoly game. Early entry is beneficial because the bank would then learn its skills in the new activity. This allows the bank to compete more aggressively when it has favorable information about its skills and more cautiously when it has poor information about its skills. The benefits of early entry also depend on how likely it is that a specialized provider will come along. Whether early entry is optimal will thus crucially depend on the competitive environment.

4.2 Importance of the competitive environment

Also the competitive environment of the existing banking activities enters the analysis because of the investment and risk associated with early entry in the new activity. If banking is sufficiently competitive, banks would be unable to absorb the investment and risk that come with early entry. An immediate implication is that investments in strategic options and thus the adoption of broader, less-focused strategies will be observed in less competitive industries, whereas firms in competitive industries will embrace more focused strategies. This could explain why Continental European banks generally follow broad strategies. Their

local market power allows them to afford the "widening of scope" strategy and benefit from its potential future strategic advantages.

Moreover, as stated earlier, the anticipated future competitive environment for the new activity matters as well. If the bank anticipates facing little or no competition in this activity in the future, early entry – with its accompanying cost of focus dilution – is unnecessary because a competitively unchallenged bank can operate successfully in this market without the benefit of early skills discovery. At the other extreme, when the anticipated competition for the new activity is very intense (perhaps due to many potential future competitors), early entry may once again be sub-optimal. The analysis in Boot, Milbourn and Thakor [1999] thus leads to the prediction that *moderate* anticipated competition in the new activity facilitates early entry. In Table 3 I have summarized the main insights.

Table 3. Optimal scope as function of the competitive environment

Anticipated competitive environment in the strategic option (new activity)	Current competitive environment in existing banking activities	
	Little competition	*High competition*
Little competition	Narrow	Narrow
Medium competition	Broad	Narrow
High competition	Narrow	Narrow

(Narrow - no early investment in new activity, Broad - early investment in new activity)

The analysis shows that the competition in the bank's current activity, the competition it anticipates in the future in a new activity, and the degree of uncertainty about future skills needed in the new activity combine to lead to predictions about early entry and hence optimal scope. Scope expansion is seen to be optimal when there is high strategic uncertainty, moderate competition expected in the new activity, and low-to-moderate competition in the existing activity.

In this context also the benefits of consolidation could be explored. Now assume that there are multiple competing banks at the outset. Consider two of these banks contemplating a merger. The question before them is whether consolidation (merging) today gives them a competitive advantage in undertaking the new activity tomorrow. Boot, Milbourn and Thakor [1999] show that the benefit of such a merger is twofold. First, merging may help create "deep pockets" making investments in strategic options more affordable. Second, merging leads to *diversification* in skills. The two banks jointly have a higher probability of having the right skills to compete in the new activity than

each has separately. While both effects may work in concert in many mergers, either effect by itself could rationalize a merger.

It should be clear that these effects have little significance in an environment without strategic uncertainty. The analysis thus predicts greater consolidation in industries with more strategic uncertainty.

4.3 Is strategic uncertainty special to financial services?

Why does this model of strategic uncertainty fit banking so well? There are at least three reasons. First, deregulation in financial services is opening doors to new activities for banks at a rate that is unprecedented since the Great Depression. Second, the swirling tides of technological and regulatory changes are generating a level of uncertainty about the skills needed to operate successfully in the future that is perhaps greater in banking than in any other industry. Lastly, banks have traditionally faced limited competition in their home markets. This has created "deep pockets" across the industry, and serves to support the broad strategies observed in banking. In particular, the combined validity of these arguments makes the model especially suited for the banking industry.[20]

The precise interpretation of the model of strategic uncertainty could also be amended to fit banking even better. In particular, we could interpret the bank's problem as the bank not knowing what combination of activities will give it a competitive edge in the future. Now we would not necessarily be talking about a bank entering new activities but possibly about the bank entering "old" activities that it traditionally chose to abstain from. Early entry, or better, choosing a wider set of activities would let the bank discover what activities optimally fit together. This interpretation would be fully consistent with the analysis in Boot, Milbourn and Thakor [1999].

[20] However, this does not limit the applicability of our model. In fact, any industry with similar characteristics to those given above – such as pharmaceuticals or telecommunications – is amenable to the interpretations and insights provided.

5 Political Considerations

I will discuss political considerations in the context of differences between the banking industry in Europe and the US In both, government interference has been quite dominant. Consolidation has been observed in Europe already for some time and in the US more recently. But is there really something to be learned from the European experience, or from the American experience for that matter? And are experiences across continents and countries comparable? In a recent paper on corporate governance, Jon Macey and I describe one of the big fallacies in comparing models of corporate governance. We say "the rhetoric about corporate governance appears to us to be divorced from reality in that two paradigmatic governance systems – the German model and the US model – are not really models at all. […] The German model is not even a model for the rest of Europe" (Boot and Macey [1998]). I feel the same in the context of discussing European or American trends. Of course, some obvious lessons can be learned. For example, the more consolidated financial sector observed in Europe gives a clear hint about what can be expected in US banking when regulatory constraints become less binding (as they have become in recent years). But what can be said more fundamentally about the *diverse* European experience? And what can be learned from the US experience?

Let me first focus on the arguable superficial common European experience as it may relate to the US Europe and the US share some similar dynamics. In particular, the relaxation of constraints on interstate banking in the US is reminiscent of the European Union banking directives liberating cross-border banking. However, immediately, a fundamental difference between US and Europe surfaces. The domestic banks in Europe were – and are – protected as domestic flagships. A fundamental belief that financial institutions should not be controlled by foreigners has (so far) almost prevented any cross-border merger.

The political dimension is at the root of this. Even in countries that do not have any direct interference by governments in banking operations and where banks are considered truly commercial enterprises (and have generally been successful, e.g. ABN AMRO and ING in The Netherlands), the political dimension is important. Central banks, ministries of finance and the banks operate in close concert. This is not very surprising: a very homogeneous group of executives is in charge of the financial sector, central bank and government ministries guaranteeing

a clear national identity of domestic institutions. In countries with explicit government involvement (e.g. France and Italy), foreign control over domestic institutions is even more unlikely unless banks become so inefficient and weak that involvement of foreigners becomes almost inevitable. To some extent this is happening. For example, in the recent bidding war for the French bank CIC, ABN AMRO was favored by some because of its excellent track record vis-à-vis competing French bidders.

The primary response to the liberating E.U. directives has so far been defensive: domestic mergers are generally encouraged to protect national interests. A case in point is Germany. Many have observed that banking in that country is surprisingly dispersed despite the powerful images of Deutsche Bank, Commerzbank and Dresdner Bank. Public policy definitely aims at protecting the interests of these powerful institutions, but the consolidation is played out mainly on the Länder-level (the separate states). Indeed, precisely at the level where the political dimension is at work. This is an important explanation for the regional consolidation in German banking.

So, wherever we look in Europe, I dare to conclude that the national flagship dimension has been of primary importance. Cross border expansion is rare and consolidation is primarily observed within national borders. For the US this gives little direction. Interstate expansion has been a driving force behind the consolidation in US banking. Politics does now seem to interfere little with interstate expansion. The political dimension in the US seems focused on the demarcations between commercial banking, investment banking and insurance. Powerful lobbies are successful in mobilizing (local) politicians and in this way have been able to obstruct major banking reform in the US Congress.

In other words, in both the US and Europe vested interests are at work. In Europe there are national authorities preserving their national flagships, in the US, powerful lobbies that seek to preserve traditional demarcations between financial institutions. These observations do not yet answer the question whether national (European) authorities are serving the interests of their constituencies when advocating national flagships. This is a different issue, and may have to be looked at in a game-theoretic context. If *other* countries are following these policies, an individual country may be well advised to follow the same policy. However, all would possibly be better off if none would follow a "national flagship policy". The ultimate success of such policy depends crucially on the efficiency of the financial institutions involved.

6 The Future and Implications for Sweden

6.1 Relevance of strategic options

Let me highlight a broader interpretation of the strategic option explanation in Boot, Milbourn and Thakor [1999] in the context of the restructuring of the European financial services industry. Bankers strongly belief that a strong position in the home market is crucial for a successful expansion in foreign markets. Generally, this seems to be the case. I will give a few examples. Belgian banks have weak foreign operations: the Belgian political situation (the split between the French and Dutch speaking regions) did *not* allow for strong domestic powerhouses. Swedish and other Scandinavian banks suffered from a financial crisis in the late eighties, early nineties inhibiting their foreign aspirations. Spanish banks started to consolidate "late". Their foreign aspirations seem limited, but some (e.g. Santander) choose to expand in the South American market (with some success). The Dutch, Swiss and – to a lesser extent – German powerhouses have strong franchises in their home markets and may well be the only Continental European banks with credible foreign aspirations.[21]

In the interpretation of the Boot, Milbourn and Thakor [1999] paper, strength in the home markets allows banks to invest in strategic options. An important one is investment banking (IB). While Continental European banks traditionally dominated the *domestic* activity in investment banking, they have had a more marginal role in IB in foreign markets and now also face severe competition in their domestic IB activity. Many of them feel that a presence in IB *might* be important for their existence as powerful banks in the future. They are willing to accept – for the moment at least – relatively low returns on those activities. The potential but uncertain vital role of these activities in the future defines them as a strategic option.

[21] I deliberately leave out the U.K. banks whose prospects are mixed, but definitely have a strong potential. One of my more favorable consolidation scenario's would involve cross-border mergers of Dutch and British institutions.

From a shareholder value maximization point of view, investing in strategic options might be desirable (if at least *potentially* sufficiently lucrative). However, how can we distinguish the "strategic option" explanation from a simple managerial entrenchment explanation? That is, managers (and governments!) may just want powerful institutions for their own sake. Distinguishing between those explanations is difficult. As the experiences of the French bank Credit Lyonnais teach us, banks that are not accountable, and even worse, operate as playground for government-appointed crownies are unlikely to follow value maximizing strategies. Growth then becomes a managerial entrenchment strategy.

Banks themselves are *ambivalent* too. The struggle of European banks in investment banking is a perfect example: while some see it as a strategic option, others (NatWest and Barclays) have retreated, albeit not really voluntarily! We see a similar ambivalence vis-à-vis insurance activities. Some think that it is perfectly complementary to commercial banking activities (e.g. to economize on the distribution network) and have embraced it – see ING and Credit Suisse-Winterthur – others choose to stay out of it (e.g. AEGON).

Nevertheless, as attested to in Section 3, I do believe that scale and scope economies are present in banking. I am tempted to subscribe to Calomiris and Karceski's [1998] notion of "client based universal banking strategies" where a bank seeks to optimally service its client base by choosing the appropriate products, services and geographical presence. Simultaneously, however, I observe that much of the consolidation in European banking is defensive. Consolidation has increased scale and scope mainly in domestic markets and facilitated local market power. Size has reached proportions that seriously questions whether anymore benefits of scale are present. And is the wider scope truly sustainable? Will it not cause dilution and loss of focus? If so, it will clearly limit the desirability of investing in strategic options. Instructive in this respect is that the operations of European universal banks in foreign markets (where they face more competition) are generally well-focused.

6.2 Is national identity important?

I would not dare to say that the national identity or ownership of financial institutions does not matter, albeit as free-market economist, I would choose to leave it open to market forces. As I have stated in Section 5, individual countries may *sometimes* be well advised to preserve national flagships. However, I am not suggesting in any way that the state should subsidize its financial institutions. Rather it should

facilitate their healthy growth and development. But favoritism should not always be excluded.

6.3 What about Sweden?

I list five categories of issues: *i*. important considerations for the Swedish financial sector, *ii*. main strengths of financial institutions, *iii*. weaknesses and concerns, *iv*. options for the future, and *v*. the value of alliances.

i. Important considerations

- An important distinction needs to be made between the viability of existing financial institutions in the Swedish or Scandinavian market and the likelihood that existing institutions will preserve their independence (and remain or become national flagships) in the future. Swedish banks seem ahead in the Nordic market. An noteworthy exception is the challenge by Den Danske Bank, the largest bank in Denmark that bought the Swedish regional bank Östgöta Enskilda Bank;
- Swedish banks have followed very diverse repositioning strategies, e.g. Svenska Handelsbanken (main focus: retail (mainly in Sweden) and some corporate and investment banking in Nordic market; recent merger with local mortgage lender Stadshypotek), FöreningsSparbanken (main focus: retail in local Swedish market; result of merger between Sparbanken Sverige and Föreningsbanken), Merita Nordbanken (main focus: pan-Nordic universal bank; result of a merger between the Swedish Nordbanken and Finland's Merita Bank), and SE-Banken (main focus: heavier in investment banking, trading and asset management; recent merger with insurer Trygg-Hansa);
- Investment banking in Nordic market is truly international with a sizable market share by foreign (non-Nordic) institutions;
- Strong Nordification of financial institutions. Little non-Nordic competition in traditional commercial banking activities. Similarly, Nordic players only focus on Nordic market;
- "Fortress Scandinavia" is to some extent being created, see also Stockholm Exchange that has chosen not to participate in the pan-European cooperation of major stock exchanges.

ii. Strengths of Swedish financial institutions

The Swedish financial institutions have recovered remarkably from the banking crisis of the early 1990's. In terms of efficiency numbers,

profitability and capitalization, they belong to the top of European financial institutions. Their strengths could be summarized as follows.

- high capitalization;
- cost efficient;
- high return on equity;
- leading in Nordic countries;
- culturally adapt, strong relationships with customers.

These strengths should not lead to complacency by Swedish financial institutions. Major challenges lie ahead.

iii. Weaknesses and main concerns

- small home market. The scale of operations that can be obtained within the Nordic market is relatively small;
- limited geographical diversification (Nordic market at best);
- (risk of) lack of true international perspective;
- main profitability comes from traditional commercial banking activities;[22]
- potential mismatch between regional orientation of banks and international orientation of the major industrial Nordic corporations;
- risk of new distribution channels. More specifically, direct banking by new entrants (possibly linked to asset management firms) could pose a threat;[23]

A general concern is that the regional orientation of Nordic banks potentially insulates them from new innovations. What I mean with this is that the lack of international presence reduces the possibility of "importing" innovations and/or practices from other markets. This is a serious concern. The experience is that head-to-head competition in some (foreign) markets has a very healthy impact on a bank's performance in its more protected domestic market.

iv. Options

- Geographic segmentation of financial services will not adhere to Nordic borders. If autonomy of Nordic financial institutions is an

[22] This actually is partially a strength too, because it also points at capitalizing on strong relationships.

[23] One should not under-estimate this risk for the Nordic financial institutions. The franchise value of Nordic institutions is virtually totally dependent on the existing branch network. The potential of new distribution channels exposes this to some risk. However, name-recognition is of some importance and new distribution channels may need to establish this as well.

objective, alliances with institutions elsewhere might be needed. This is particularly true for universal banks with aspirations beyond retail and the lower and middle corporate market. See also my comments below on investment banking;

- A viable role in investment banking requires "deep pockets" and an international, rather than regional presence. Neither condition is or can be met by the Swedish institutions on their own. True aspirations in that direction can only be met with mergers beyond the Nordic market;
- Narrower strategies (e.g. the "traditional" Citibank strategy: retail, asset management and lower/middle corporate market) might allow for a sustainable role as independent regional player. Focus should be on exploiting relationships and services that are targeted to this relationship-oriented clientele;
- The previous point also implies that financial institutions should aim at optimally exploiting scope economies. I doubt that this allows for broad universal banking strategies by Nordic players unless they engage in alliances, see issue/item v;
- The previous points imply that only narrower strategies are viable for independent Nordic banks. An exception is possibly asset management. This is an important market and at least one of the Nordic players might be able to capitalize on it. The combination of asset management on a large, possibly international scale and regionally focused banking might be feasible. In this way, some synergies could also be obtained with life insurance activities. That is, life insurance could fit with the relationship orientation of the regional bank and simultaneously add to the asset management pool. An alternative is "alliance model" for asset management activities, see issue/item v;
- My impression is that the main *strategic options* for Nordic banks involve expansion beyond the Nordic market (e.g. Baltic countries) and alliances for investment banking and/or asset management. For some, real integration of life insurance and (possibly) pension business may add directly to the existing relationship-oriented banking activities.

v. Value of alliances

The concept of alliances has yet to be developed in the context of banking. This is to some extent surprising. Banks did, and still do, engage in correspondent banking, particularly in the context of cross-border payment services. But correspondent banking is losing its importance. Why? With the advent of information technology international payment and settlement systems have become available

(e.g. the emergence of TARGET and settlement systems like Cedel and Euroclear). These developments reduce the need for correspondent banking. More importantly, correspondent banks may have become competitors in the areas they were cooperating in before. For example, some banks seek to gain a competitive edge by offering proprietary cross border payment facilities. This points at an important consideration for the feasibility of correspondent banking, or alliances for that matter. It only works if the interests of the participating institutions are sufficiently aligned.[24] But why may alliances become important, particularly, within the context of Nordic banking?

The main reason I see is that institutions that seek to capitalize on their local presence, with benefits rooted in strong relationships, cultural adaptation, etc., may need smooth access to some investment banking and asset management services that are scale intensive and globally, rather than locally oriented. It may well be possible to offer some of these services in an alliance (i.e. "to join forces") and capitalize on customer-related synergies. In the context of Nordic banking, I could envision cooperation along this dimension with (regional) banks outside the Nordic countries. While some will argue that a merger with these institutions would allow for a smoother operation of these services, I would like to take issue with this point of view.

First, for several reasons, cross border mergers may not (yet) be feasible. A focused alliance would create valuable linkages between institutions with immediate synergy benefits (see above), but could also allow the possibly nationally-rooted partners to "get to know" each other. In that sense, it would be an intermediate phase. As a second argument, the alliance-model based on asset management and/or specific investment banking activities may, if properly designed, combine the benefits of an integrated universal banking structure and a stand-alone type of organization of those activities. For example, the alliance partners all have a limited exposure to these activities which helps them maintain focus. In particular, "cultural" conflicts and distractions associated with trying to build up (or buy) an investment bank next to running the relationship-rooted regional bank are prevented.[25] Obviously, the alliance model does not come without cost. The important task is to define a clearly defined portfolio of activities that would become part of the alliance. This will not be investment banking in the broadest sense of

[24] Observe that correspondent banks could traditionally not enter each others markets. Interests were therefore more readily aligned.

[25] The experience of some western banks is that top management gets fully distracted by the investment banking activities and spends disproportionally little time on the often more profitable non-investment banking activities.

the word. Similarly, in the case of asset management, the alliance partners would each maintain their own proprietary access to the customers but join forces in the asset management operations including research and back office activities. This would facilitate the information technology investments that allow the partners to capitalize on scale economies. Maintaining proprietary access by the individual alliance partners preserves customer-related scope economies.

6.4 The future

There are powerful forces behind consolidation. I believe that consolidation is only partially driven by value-maximizing behavior. As I have emphasized, also the political dimension cannot be ignored. Consolidation in Europe and the US will continue. The regional expansion that characterizes much of the US merger wave will carry over to Europe. Cross-border acquisitions are coming, particularly with the arrival of the Euro and the European Monetary Union (EMU). The Euro and EMU are catalysts and will accelerate the integration of national financial markets, and induce a more pan-European view on financial services.[26]

The merger wave will continue and, in my view, become excessive (overshooting). Ultimately, it will lead to a level of consolidation in the industry that will be reversed and lead to downsizing and refocusing. In the consolidation phase, however, larger financial institutions will be "forced" to follow "the herd". Swedish banks have rightfully chosen to expand in the Scandinavian territory. This type of regional expansion is in my view value enhancing. Whether Scandinavian banks will ultimately survive as independent financial institutions is questionable. In my view, alliances are crucial to maintain the viability of regionally focused banks. But even then, once the regional expansion in the Scandinavian market is completed, the surviving institutions become prime targets for the powerful Continental European institutions.

[26] I have said little about moral hazard and regulation in light of the consolidation. For one reason there is broad agreement: certifying risk management processes is the primary task of supervision, and the EU has quite wisely allocated this task unambiguously to home country supervisors. This is something still to be addressed in the U.S. observing the multiplicity of regulators. The consolidation, and broadening of scope now also observed in the U.S. (e.g. Travelers and Citicorp), amplifies the importance of this issue.

References

Allen, F., (1993), ”Stock Markets and Resource Allocation”, in *Capital Markets and Financial Intermediation*, C. Mayer and X. Vives, eds., Cambridge University Press.

Allen, F. and D. Gale, (1995), ”A Welfare Comparison of the German and US Financial Systems”, *European Economic Review*, 39, 179-209.

Berger, A.N., A.K. Kashyap and J.M. Scalise, (1995), ”The Transformation of the US Banking Industry: What a Long, Strange Trip It's Been”, *Brookings Papers on Economic Activity*, 2, 55-218.

Berger, A.N., (1997), ”The Efficiency Effects of Bank Mergers and Acquisition: A Preliminary Look at the 1990s Data, forthcoming”, in *Bank Mergers and Acquisitions*, Yakov Amihud and Geoffrey Miller, eds, Klumer Academic, Boston, MA.

Berger, A.N., R.S. Demsetz and P.E. Strahan, (1999), ”The Consolidation of the Financial Services Industry: Causes, Consequences and Implications for the Future”, *Journal of Banking and Finance*, 23, 135-194.

Berger, A.N., W.C. Hunter, and S.G. Timme, (1993), ”The Efficiency of Financial Institutions: A Review and Preview of Research Past, Present and Future”, *Journal of Banking and Finance*, 17 (special issue on *The Efficiency of Financial Institutions)*, 221-249.

Berger, P.G. and E. Ofek, (1995), ”Diversification’s Effect on Firm Value”, *Journal of Financial Economics*, 37, 39-65.

Berglöf, E. and E.L. von Thadden, (1994), ”Short-Term versus Long-Term Interests: Capital Structure with Multiple Investors”, *Quarterly Journal of Economics*, 109, 1055-1084.

Berlin, M., (1996), ”For Better and for Worse: Three Lending Relationships”, *Business Review Federal Reserve Bank of Philadelphia*, December, 3-12.

Bhattacharaya, S. and G. Chiesa, (1995), ”Proprietary Information, Financial Intermediation, and Research Incentives”, *Journal of Financial Intermediation*, 4, pp. 328-357.

Boot, A.W.A., S.I. Greenbaum and A.V. Thakor, (1993), ”Reputation and Discretion in Financial Contracting”, *American Economic Review*, 83, 1165-1183.

Boot, A.W.A. and S.I. Greenbaum, (1995), ”The Future of Banking: What Should Corporate America Expect?”, *Business Week Executive Briefing*, 8.

Boot, A.W.A. and J. Macey, (1998), “Objectivity, Proximity and Adaptability in Corporate Governance”, working paper, University of Amsterdam.

Boot, A.W.A., T. Milbourn and A.V. Thakor, (1999), ”Evolution of Organizational Scale and Scope: Does It Ever Pay to Get Bigger and Less Focused?”, working paper, University of Amsterdam.

Boot, A.W.A. and A. V. Thakor, (1997), ”Financial System Architecture”, *Review of Financial Studies*, 10, 693-733.

Boot, A.W.A. and A.V. Thakor, (1999), ”Can Relationship Banking Survive Competition?”, working paper, University of Amsterdam.

Boot, A.W.A., A.V. Thakor and G. Udell, (1991), ”Credible Commitments, Contract Enforcement Problems and Banks: Intermediation as Credibility Assurance”, *Journal of Banking and Finance*, 15, 605-632.

Boyd, J.H. and M. Gertler, (1994), ”Are Banks Dead, or Are the Reports Greatly Exaggerated?”, working paper, Federal Reserve Bank of Minneapolis.

Calomiris, C.W. and J. Karceski, (1998), “Is the Bank Merger Wave of the 90’s Efficient? Lessons from nine Case Studies”, working paper, Columbia University.

Canals, J., (1994), “Competitive Strategies in European Banking”, Clarendo Press, Oxford.

Chan, Y., S.I. Greenbaum and A.V. Thakor, (1986), ”Information Reusability, Competition and Bank Asset Quality”, *Journal of Banking and Finance*, 10, 255-276.

Clark, J.A., (1996), ”Economic Cost, Scale Efficiency, and Competitive Viability in Banking”, *Journal of Money, Credit and Banking*, 28, 342-364.

Comment, R. and G.A. Jarrell, (1995), Corporate Focus and Stock Returns, *Journal of Financial Economics*, 37, 67-87.

Cornett, M.M. and H. Tehranian, (1992), ”Changes in Corporate Performance Associated with Bank Acquisition”, *Journal of Financial Economics*, 31, 211-34.

Courtney, H., J. Kirkland and P. Viguerie, (1997), ”Strategy under Uncertainty”, *Harvard Business Review*, November-December, 67-79.

Dewatripont, M. and E. Maskin, (1995), ”Credit and Efficiency in Centralized and Decentralized Economies”, *Review of Economic Studies*, 62, 541-555.

Dewatripont, M. and J. Tirole, (1995), *The Prudential Regulation of Banks*, MIT Press, Cambridge, Massachussets.

Diamond, D., (1984), ”Financial Intermediation and Delegated Monitoring”, *Review of Economic Studies*, 51, 393-414.

Diamond, D., (1993), ”Seniority and Maturity of Debt Contracts”, *Journal of Financial Economics*, 33, 341-368.

Flannery, M., (1999), ”Comment on Milbourn, Boot and Thakor”, *Journal of Banking and Finance*, 23, 215-220.

Gorton, G. and J. Kahn, (1993), ”The Design of Bank Loan Contracts, Collateral, and Renegotiation”, NBER working paper 4273.

Gorton, G. and G. Pennacchi, (1995), ”Banks and Loan Sales: Marketing Nonmarketable Assets”, *Journal of Monetary Economics*, 35, 389-411.

Greenbaum, S.I. and A.V. Thakor, (1995), *Contemporary Financial Intermediation*, Dryden Press.

Grossman, S.J. and O. Hart, (1986), ”The Costs and Benefits of Ownership: A Theory of Vertical and lateral Integration”, *Journal of Political Economy*, 94-4, 691-719.

Hamel, G. and C.K. Prahalad, (1990), ”The Core Competence of the Corporation”, *Harvard Business Review*, May-June, 79-91.

Hellwig, M., (1991), ”Banking, Financial Intermediation and Corporate Finance”, in *European Financial Integration*, A. Giovanni and C. Mayer, eds., Cambridge University Press.

Houston, J. and C. James, (1995), Bank Information Monopolies and the Mix of Private and Public Debt Claims, working paper, University of Florida.

James, C., (1987), ”Some Evidence on the Uniqueness of Bank Loans”, *Journal of Financial Economics*, 19, 217-235.

John, K. and E. Ofek, (1995), ”Asset Sales and Increase in Focus”, *Journal of Financial Economics*, 37, 105-126.

Kroszner, R.S. and R. Rajan, (1994), ”Is the Glass-Steagall Act Justified? A Study of the US Experience with Universal Banking Before 1933”, *American Economic Review*, 84, 810-832.

Kulatilaka, N. and E. Perotti, (1998), ”Strategic Growth Options”, *Management Science*, 44 (8), 1021-1031.

Lummer, S. and J. McConnell, (1989), ”Further Evidence on the Bank Lending Process and the Reaction of the Capital Market to Bank Loan Agreements”, *Journal of Financial Economics*, 25, 99-122.

Mayer, C., (1988), ”New Issues in Corporate Finance”, *European Economic Review*, 32, 1167-1183.

Merton, R.C., (1993), ”Operation and Regulation in Financial Intermediation: A Functional Perspective”, in *Operation and Regulation of Financial Markets*, P. Englund, ed., Economic Council.

Mester, L.J., (1992), ”Traditional and Nontraditional Banking: An Information-Theoretic Approach”, *Journal of Banking and Finance*, 16, 545-566.

Mitchell, K. and N.M. Onvural, (1996), ”Economies of Scale and Scope at Large Commercial Banks: Evidence from the Fourier Flexible Functional Form”, *Journal of Money, Credit, and Banking*, 28, 178-199.

Petersen, M. and R. Rajan, (1994), ”The Benefits of Lending Relationships: Evidence from Small Business Data”, *Journal of Finance*, 49, 1367-1400.

Petersen, M. and R. Rajan, (1995), ”The Effect of Credit Market Competition on Lending Relationships”, *Quarterly Journal of Economics*, 110, 407-443.

Puri, M., (1996), ”Commercial Banks in Investment Banking: conflict of Interest or Certification Role?”, *Journal of Financial Economics*, 40, 373-401.

Rajan, R., (1992), ”Insiders and Outsiders: The Choice Between Informed and Arm's Length Debt”, *Journal of Finance*, 47, 1367-1400.

Sharpe, S., (1990), ”Asymmetric Information, Bank Lending, and Implicit Contracts: A Stylized Model of Customer Relationships”, *Journal of Finance*, 45, 1069-1087.

Schmeits, A., (1997), Discretion in Bank Contracts and the Firm's Funding Source Choice between Bank and Financial Market Financing, working paper, Washington University.

Shaffer, S. and E. David, (1991), Economies of Superscale in Commercial Banking, *Applied Economics*, 23, 283-293.

Simon, H.C., (1936), Rules versus Authorities in Monetary Policy, *Journal of Political Economy*, 44, 1-30.

Swedish Banking Efficiency and Productivity in an International Perspective

by

Lennart Hjalmarsson,*

Irene Andersson and

Aziz Mlima

Supplement number 28 to the Government Inquiry on the International Competitiveness of the Swedish Financial Sector.

Views, suggestions, and opinions expressed in this study represent the personal views of the authors and not necessarily those of the Government Inquiry on the International Competitiveness of the Swedish Financial Sector.

September, 1999

* Professor i Nationalekonomi vid Handelshögskolan, Göteborgs Universitet. Hjalmarsson forskar inom områdena produktivitet och effektivitet samt energiekonomi och avreglering av marknader. Department of Economics, Göteborg University, P.O.Box 640, SE-405 30 Göteborg, Sweden.
E-mail: Lennart.Hjalmarsson@economics.gu.se

Table of Contents

Sammanfattning (in Swedish)

Syftet med denna rapport är att analysera den relativa effektiviteten och produktivitetsutvecklingen inom svensk bankverksamhet. Rapporten inehåller dels en redovisning av ett antal tidigare studier av svensk bankverksamhet, dels vissa nya resultat. På metodsidan utnyttjas olika typer av frontfunktionsanalys, speciellt s.k. stokastiska frontfunktioner samt DEA (Data Envelopment Analysis).

Det stora problemet vid mätning av banksektorns produktivitet är hur produktionen ska definieras och mätas. Utgör inlåningskonton input eller output i bankverksamheten? Mäts banktjänster bäst genom antalet konton och transaktioner eller värdet på dessa? Är hanterandet av konton verkligen den mest betydelsefulla prestationen i bank-verksamhet? Är det inte istället bankernas roll som förmedlare av kredit från sparare till investerare som, samhällsekonomiskt sett, är den viktigaste prestationen? Och hur ska i så fall denna mätas?

Vid analys av bankers produktion är det två huvudansatser som dominerar i litteraturen med avseende på vilka aspekter som studeras:

- produktionsansats (tjänste- eller förädlingsvärdeansats)
- förmedlingsansats (tillgångsansats)

I produktionsansatsen betraktas banker som producenter av inlåning, utlåning och andra tjänster med hjälp av arbetskraft och kapital. I denna ansats utgör antalet inlåningskonton, antal utestående lån, antal valutatransaktioner etc. outputvariabler. I förmedlingsansatsen betraktas banker som hopsamlare av kapital som sedan omvandlas och förmedlas till lån och andra tillgångar. Volymen i kronor på inlåningskonton och utlåning utgör här output och rörelsekostnader och räntor input i produktionsprocessen. Någon exakt klar gräns mellan de två ansatserna existerar dock inte, utan valet av variabler bestäms nog oftare av tillgängligheten snarare än av ansatsen.

En viktig fråga är hur skillnader i effektivitet mellan produktionsenheter ska tolkas. Det är då viktigt att notera att effektivitet som det här definierats är ett mått på avståndet från en enhet till en best-practice front och inte ett mått på någon ”högre form” av effektivitet. Det är också viktigt att notera att vi här endast beaktar resursinsats och produktion i mer eller mindre fysiska termer. Effekter av varierande produktpriser eller faktorpriser, lönsamhet eller avkastning ligger utanför analysen. Med vissa reservationer kan vi säga att en effektivitetsanalys

kan ge en indikation på enheternas relativa konkurrenskraft på en konkurrensutsatt marknad. Detta innebär att inom starkt reglerade sektorer eller sektorer med svag konkurrens kan korrelationen mellan effektivitet och lönsamhet vara mycket svag. Goda finansiella nyckeltal kan lika gärna vara ett uttryck för monopol-makt som för effektivitet

Rapportens huvudresultat kan sammanfattas på följande sätt:

- Det existerar ingen generell samsyn i litteraturen när det gäller definitionen av vad som ska klassificeras som producerade tjänster i bankverksamhet och vad som är att betrakta som insatsfaktorer. Det finns en stor variation mellan olika studier. Även inom de två dominerande ansatserna, produktionsansatsen och förmedlingsansatsen finns det en betydande variation i val av tjänsteproduktionsvariabler och insatsfaktorsvariabler. Eftersom de empiriska resultaten inte tycks vara speciellt robusta när det gäller val av variabler på output- och inputsidan utgör naturligtvis detta ett problem vid tolkningen av de empiriska resultaten och vid jämförelser mellan olika undersökningar.
- Ett generellt resultat när det gäller bankeffektivitet och produktivitetsförändringar är att resultaten tycks vara mycket känsliga för variationer i servicevolymerna, dvs. kapacitets-utnyttjandet. I de studier som uppmärksammas i denna rapport gäller det speciellt för produktivitetsförändringar på bankkontors-nivån, där produktiviteten varierar mycket starkt mellan åren, väl korrelerad med variationer i tjänstevolymen. Detta visar sig också vara fallet under åren efter bankkrisen. Produktivitetsutvecklingen på fronten övergår i en produktivitetstillbakagång efter 1992. A priori skulle man förvänta sig en kraftig rationaliseringsfas med snabb produktivitetsökning. Det som förklarar resultaten är istället den anpassning av hushållens och företagens portföljer som blev resultatet av den kraftiga ökningen i realräntan med omfattande återbetalning av lån och lägre efterfrågan på nya lån. Bankerna hade uppenbarligen inte möjlighet att anpassa arbetskraft och andra insatsfaktorer i proportion till förändringarna på tjänstevolymsidan.
- Effektivitetsanalysen av de svenska bankerna 1996 illustrerar resultatens känslighet med avseende på valet av variabler. Å ena sidan visar sig valet mellan arbetskraftskostnader och arbetstimmar ha endast obetydlig effekt på effektivitetsfördelningen. Å andra sidan visar sig resultaten vara mycket känsliga för inlemmandet eller utelämnandet av nettoprovisioner som en tjänstevariabel. Medan resultaten för de kommersiella bankerna inte påverkas nämnvärt har denna variabel mycket stor effekt på ett antal sparbankers effektivitetsmått.

- I ett skandinaviskt perspektiv var den svenska banksektorn 1990 betydligt mera produktiv än såväl den finska som den norska och danska banksektorn. Denna slutsats gäller också för de största bankerna. Det stora gapet mellan svensk och finsk genomsnittsproduktivitet beror i huvudsak på effektivitetskomponenten, även om Sverige också hade ett teknologiskt försteg.
- Vid en jämförelse mellan ett stort antal europeiska affärsbanker visar sig den genomsnittliga effektivitetsnivån för de svenska affärsbankerna ligga på andra plats från botten, både i en tvåfaktormodell och i en arbetseffektivitetsmodell. Det är då viktigt att här komma ihåg att analysen är begränsad endast till ett antal affärsbanker i varje land samt till ett fåtal variabler.
- I allmänhet visar det sig vara mycket stor variation i effektivitet mellan olika banker både inom och mellan länder. Jämfört med många andra branscher kan man också à priori förvänta sig en lägre effektivitet och en större spridning i effektivitet i banksektorn och ett större avstånd mellan de mest effektiva och de minst effektiva bankerna. En viktig orsak till detta är det relativt låga konkurrenstrycket i banksektorn i kombination med en blandning av banker med olika typer av ägare och målsättningar, affärsbanker, kooperativa banker och stiftelser utan vinstsyfte, börsnoterade och icke börsnoterade företag. Närliggande orsaker är också den regleringsstruktur som bankerna möter i olika länder och som ytterligare kan öka avståndet mellan de mest effektiva och minst effektiva enheterna.
- Vid en bedömning av den svenska banksektorns internationella konkurrenskraft kan konstateras att samtliga länder uppvisar en stor spridning mellan de mest effektiva och de minst effektiva bankerna. Å andra sidan tyder resultaten på att den svenska banksektorns genomsnittsproduktivitet är relativt låg jämfört med övriga EU-länders. I ett nordiskt perspektiv tycks dock den svenska banksektorns produktivitet klart överstiga såväl den finska som den norska och i vissa fall även den danska.

Summary

In this report we have briefly surveyed and to some extent extended a couple of empirical studies on Swedish banking efficiency and productivity change on the basis of frontier production function methods. The results may be summarised in the following way:

- In the literature, there is no general agreement on the definition of outputs and inputs in banking, and there is a lot of variation between different studies. Even within each of the two dominant approaches, the production approach (or service provision or value added approach) and the intermediation approach (or asset approach), there is substantial variation Moreover, the empirical results do not seem very robust, but rather depend upon the definition of outputs and inputs – although rather few studies have dealt with this issue.
- In general, banking efficiency and productivity change seem to be very sensitive to fluctuations in output variables, i.e., in capacity utilisation. This is most evident in the case of productivity change at branch offices, which shows high variation between years, strongly correlated with output variation. This is also evident in the years after the banking crisis, when technical change at the frontier turns into regress, from 1992 onwards. As a response to a banking crisis, this may seem counter intuitive. However, it turns out that the result, to a large extent, is driven by the development of output. A major reason for the banking crisis was the dramatic increase in the real rate of interest, causing large changes in household and company portfolios, with extensive repayment of loans and less demand for new loans. While deposits decreased somewhat, loans and guarantees decreased substantially, and it was obviously not possible to adjust labour use at the same rate.
- The efficiency analysis of Swedish banks in 1996 illustrates the sensitivity of the results with respect to choice of variables. On the one hand, the choice between labour costs and labour hours has only minor impact on the efficiency distribution. On the other hand, the results are very sensitive to the inclusion or exclusion of net provisions as an output variable. While the results are robust for the commercial banks – which are all (with one exception) fully efficient, the deviation between models is striking for some savings banks.
- In a Scandinavian comparison, the Swedish banking sector in 1990 was much more productive than the Finnish, Norwegian, and Danish

banking sectors. This conclusion also holds for the largest banks. The large gap between Swedish and Finnish average productivity is mainly due to the efficiency component, although Sweden also has a technology advantage.

- In a European comparison, Swedish commercial-banking efficiency ranks second from the bottom, both in a two-input model and in a labour use model. Important caveats here are the small sample of banks included in the study, and the small number of inputs and outputs covered.
- In general, there is large variation in efficiency between banks both within each country analysed and across countries. In comparison with many other industries, one should also expect lower mean efficiency and a larger dispersion in efficiency scores, i.e., between best practice and worst practice in the banking industry. The main reason for this is the rather low competitive pressure in the banking industry in combination with a mixture of banks with different types of ownership and objectives, commercial banks, cooperative banks and not-for-profit savings banks. A related reason is the different regulatory structures facing banks in different countries, which may further increase the distance between best practice and worst practice in international comparisons.
- Concerning the international competitiveness of Swedish banking, we thus observe a large variation between best practice and worst practice banks in all countries and across countries. In a European comparison the results suggest the average productivity of Swedish banks to be rather low, while in a Scandinavian comparison, we get the opposite conclusion.

Abstract

The purpose of this report is to analyse relative efficiency and productivity growth in the Swedish banking sector. We will survey some previous results for the Swedish banking industry, in addition to presenting some new results. One analysis focuses on productivity change at branch-office level. We are also investigating the impact on productive efficiency and productivity growth of the deregulation of the Swedish banking industry in the mid-1980s and the consequent banking crisis. The distribution of productive efficiency for Swedish banks in 1996 is analysed on the basis of a third data set. The final analysis compares the efficiency of large Swedish banks internationally.

1 Introduction

The purpose of this report is to analyze relative efficiency and productivity growth in the Swedish banking sector. In recent years, relative efficiency and productivity change in banking have been the focus of a lively debate, though there is still a lack of clarity as to its actual development. The banking sector has also received considerable attention in international studies. Traditionally, the areas of main interest have been analyses of economies of scale and economies of scope, though more recently a number of efficiency and productivity studies have been undertaken. In the aftermath of liberalisation, restructuring, and deregulation of financial markets around the world, there is a recent boom, in activities directed to evaluating the performance of financial institutions. In a recent paper, Berger and Humphrey (1997) survey and contrast the results of 130 financial-institution efficiency studies. In addition to the many international studies, there are also a few studies based on data for Swedish banks; see Andersson (1999), Battese et al (1997), Berg et al (1993), Heshmati (1997), Hjalmarsson et al (1991), Hoover (1998), and Kumbhakar et al (1998). In this paper we will summarise some previous results for the Swedish banking industry, in addition to presenting some new ones.

A major problem related to the measurement of banking sector productivity is how to define, let alone measure, the sector's output. Triplett (1992), pointed out that progress in measurement of banking has been inhibited by two major unresolved but related questions: (1) What are the outputs? (2) What are the inputs? For instance, are deposit accounts inputs or outputs in its activity? This ambiguity of inputs and outputs was also highlighted by Wykoff (1992): ”When deposits are outputs, why are they so cheap? When they are inputs, why do people provide them to banks?” Are bank services best measured by the number of accounts held and transactions executed, or should the value of transactions be used instead? Is the administration of these accounts really the most important aspect of banking? Is not the bank's role as intermediary between savers and investors the most important from the economic point of view? How would this be measured? For a thorough theoretical and conceptual discussion of these issues, see Berger and Humphrey (1992), Fixler and Zieschang (1992) and Colwell and Davis (1992).

Even in 'practical' work, there is an intense discussion relating to the measurement and valuation of a bank's output. Valuing bank production in national accounts has always been controversial. Given the guidelines used today, the banking and financial sector has only a weak contribution (positive or negative) to the growth of GDP. In relation to the revised international system of national accounts (SNA), a number of alternatives regarding the treatment of banking and financial activities have been suggested. In most cases, the suggestions would imply a marked increase in the sector's contribution to GDP.

Humphrey (1984) divided studies of banks' production into two distinct strands, depending on the nature of the analytical approach:

- the production approach, also called the service provision or value added approach
- the intermediation approach, also called the asset approach

In the production approach, banks provide services to customers by handling their financial transactions, keeping customers deposits, issuing loans, cashing cheques, etc. Here the number of deposit accounts, outstanding loans, foreign exchange transactions, etc., are the output variables, with labour and capital as inputs

In the intermediation approach, banks are performing the two major roles of mobilising and distributing resources efficiently, in order to smooth investment activities in the economy. In this approach, the inputs are labour, materials, and deposits, while outputs are loans and other income generating activities to the banks, see e.g. Mester (1997). In some studies, the monetary volume of the deposit accounts and the loans comprise the output, while the costs of operation and the interest due are the inputs in the production process. This approach has two major sub-groups: (a) the *profit approach*, and (b) the *risk management approach.*

In the profit approach, the bank manager’s target is to optimise the use of resources to maximise the bank’s profit function. In this approach, the bank manager cares for all types of costs in the banks and the output is the net income generated after incurring certain costs in the production process.

In the risk management approach, banks transform assets, on the basis of evaluation of the risks attached to various forms of asset. In risk management, banks take some risks to produce acceptable returns. A bank's performance will affect its valuation in the market, its ability to acquire other banks or to be acquired at a good price, and its ability to be funded through deposits or financial markets.

In both the production approach and the intermediation approach, the definition of output is strictly limited to production within banking. To what extent banks actually fulfill their role as producers of financial

services lies beyond the scope of this analytical framework. This limitation is acceptable if the intention, as here, is to estimate productivity growth separately for different sub-sectors of the economy. Spillover effects between sectors would then manifest themselves as higher (or lower) productivity in those sectors which are exposed to these services. Increased productivity in the communications network, for exampel could be expected to spill over into higher productivity within transport-dependent sectors, while well-performing banks could lead to higher productivity in the whole economy, etc. These external effects could be important but, nevertheless, fall outside the scope of a purely sectoral productivity analysis.

This report contains several pieces of empirical research related to Swedish banks. For a start, we will survey (and slightly develop) a few studies of efficiency and productivity change in Swedish banking. Then we will focus on the efficiency and productivity of Swedish banking in international comparisons.

In one part of this report the production approach is used to review how one of the largest Swedish commercial banks use labour and capital to produce certain well-defined services at branch offices. However, we are not in a position to make any judgements on their quality. Instead, we will study only the development of productivity in services, under the assumption that their quality remains unchanged over the sample period.[1]

In other parts of this report, the intermediation approach is used. In these cases the sample is collected from banks´ annual reports. Here micro units are entire banks, with one or more offices.

Because this report focuses on productive efficiency and productivity change, we use various kinds of frontier models, described in Section 2. The empirical results are presented in Section 3.

[1] Other studies with a similar approach are Sherman and Gold (1985), Tulkens (1993) and Vassiloglou and Giokas (1990). In these studies, the analysis is based on the spread of efficiency between bank offices for a particular year.

2 Measurement of efficiency and productivity change

2.1 Efficiency

Measurement of efficiency in the sense of evaluating the performance of production units against a benchmark or yardstick, under the heading of frontier analysis, is becoming increasingly widespread. Frontier analysis provides a ranking of production units in terms of their efficiency, and it also attributes a numerical efficiency value to each production unit. This makes it possible to identify *best practice* and *worst practice* in a sector, and the measurement of efficiency has in fact been the main motivation for the intensive study of frontier functions.

Thus, for a variety of reasons, in recent years frontier analysis has become a very attractive and rather sophisticated way to benchmark the relative performance of production units. First, the notion of a frontier is consistent with the underlying economic theory of optimising behavior. Second, deviations from a frontier have a natural interpretation as a measure of the efficiency with which economic units pursue their technical or behavioral objectives. Finally, information about the structure of the frontier, and about the relative efficiency of economic units has many policy applications. One example is the setting of productivity targets in price-cap regulation of natural monopolies like electricity distribution.

To serve as a meaningful benchmark, the frontier should represent best-practice technology or behaviour. This benchmark may be profits, costs or production. Estimation or construction of *profit frontiers*, *cost frontiers,* or *production frontiers* can be done in various ways. But since there is not enough engineering information available on the technology of banking, and since organisation is also of great importance, the frontier itself represents observed best practice. i. e., studies of banking efficiency rely on accounting measures of outputs, inputs, costs, revenues, profits, etc. However, there are still several competing methods for determining the best-practice frontier. These methods differ primarily in the assumptions imposed on the data in terms of:

- Functional form of the frontier function, especially the choice between a parametric form and a non-parametric one.
- Stochastic properties, especially the choice between a deterministic and a stochastic approach.
- Distribution assumed for the inefficiency scores (usually half-normal, truncated normal, exponential, or gamma distribution).

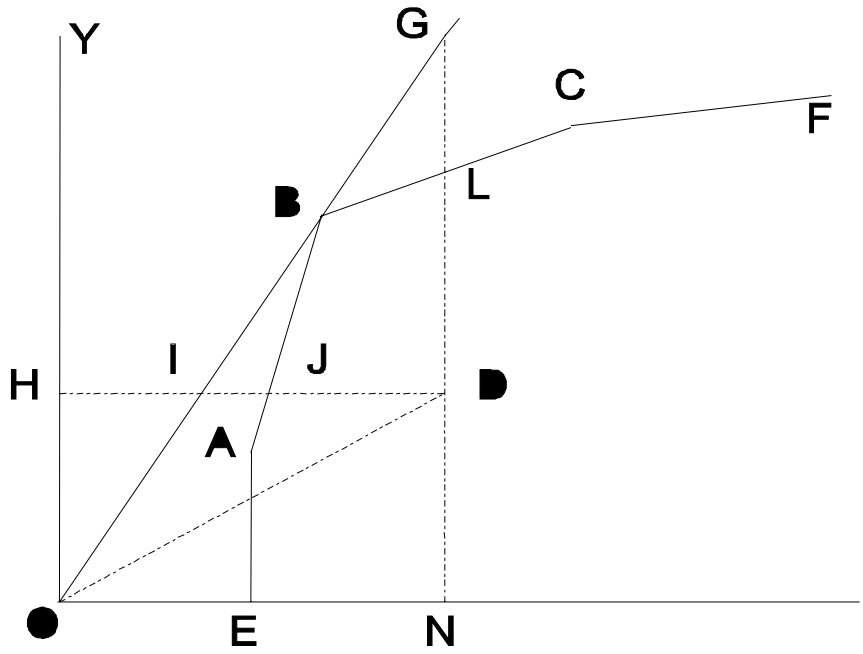
In this report we are only concerned with production frontiers; i.e., our data consists of output and input variables. In measuring efficiency we adopt the system introduced by Forsund and Hjalmarsson (1974), a generalisation of Farrell's measures to a variable-returns-to-scale (VRS) technology[2]. The generalised Farrell measures are illustrated in Figure 1, where a one – input (X) and one – output (Y) production activity is assumed. The production units are denoted by A, B, C and D. In the case of constant returns to scale (CRS), the frontier is the straight line through B and only unit B is on the frontier; i.e., only unit B is a fully efficient best-practice unit, while units A, C, and D are inefficient. With the more flexible variable returns to scale technology, the frontier is supported by three units, A, B, and C; i.e., these units are fully efficient, while D is still inefficient.

A measure of efficiency is the distance between an inefficient unit and the frontier, which can be measured in several ways. Here we focus on input-saving (or input-oriented) efficiency. Another option is output-increasing (or output-oriented) efficiency. In the case of CRS, these measures coincide; see Forsund and Hjalmarsson (1974 and 1979b). The input-saving measure shows what proportion of observed input would have been necessary to produce observed output, if the unit in question were on the efficient frontier. Similarly, the output-increasing measure shows the ratio between observed output and how much output could

[2] The basic idea of measuring efficiency on the basis of frontiers goes back to Farrell (1957), who introduced a deterministic non-parametric programming approach. The next step was taken by Aigner and Chu (1968), who introduced the deterministic parametric approach, which was then generalised by Forsund and Hjalmarsson (1979a). At the same time, Charnes, Cooper, and Rhodes (1978) developed data envelopment analysis, DEA, as a much more convenient way to calculate Farrell efficiency measures on the basis of a non-paramtreic programming approach. Econometric estimation of frontier functions was introduced by Richmond (1974), although the most popular so-called *composed error model* appeared in two independent papers by Aigner et al (1977) and Meeusen and van den Broeck (1977); for recent surveys, see Greene (1993), Lovell (1993), and Grosskopf (1996).

have been produced, with observed input and frontier technology. Efficient frontier units have a value of 1, inefficient ones less than 1.

Figure 1. An illustration of efficiency measures



VRS efficiency measures for any unit at D are:
E1 = HJ/HD = input-saving technical efficiency
E2 = ND/NL = output-increasing technical efficiency
The CRS efficiency measure is
E = HI/HD = ND/NG = technical efficiency

Unit D will be both input-saving and output-increasing inefficient under VRS, since HJ/HD < 1 and ND/NL < 1.

There are several summary measures of the efficiency of the entire sector, so called structural efficiency measures. One such measure is the mean of efficiency scores, or a size-weighted mean of the efficiency scores. Another is based on the average unit (an arithmetic or geometric mean of the industry), for which input-saving or output-increasing efficiency measures can be calculated.

As regards determination of the frontier, two approaches dominate today:

- SFA = Stochastic frontier analysis – the dominant econometric approach

- DEA = Data envelopment analysis – the dominant programming approach

SFA uses a parametric representation of technology, along with a two-part composed error term. One part of the composed error term represents statistical noise, and is assumed to follow a normal distribution. The other part represents technical inefficiency, and is usually assumed to follow a skewed distribution such as a half normal, truncated normal, exponential, or gamma distribution. One problem with cross-sectional data in efficiency measurement is that technical inefficiency cannot be separated from firm-specific effects, which may not be related to inefficiency. This problem can be avoided if panel data is available. (By panel data, we mean data on a cross-section of firms, each observed for a number of time periods.) Then the residual may be split into four components; a firm-specific effect, a time-specific effect, an inefficiency component, and random noise. Often the time specific effect is replaced by the explicit introduction of technical change in the production function.

One weakness with SFA is that if the functional form is mis-specified, measured efficiency may be confounded with specification errors. Another weakness of this approach is its limited usefulness in the case of multiple output – multiple input production when input price data is not available. A third weakness is its sensitivity to heteroscedasticity; i.e., the impact on the estimated efficiency scores when firms in a cross-section vary widely in size. If this is not accounted for, one would expect overestimation of the intercept of the frontier function and underestimation of the slope coefficients; see Caudill and Ford (1993).

The major competing frontier approach is the very popular non-parametric data envelopment analysis (DEA) approach. This approach imposes less structure on the frontier, which is formed as the piecewise linear combination of best-practice units, yielding a convex production possibilities set. Moreover, it is easy to apply in the case of multiple output and multiple input. It involves the construction of a technological frontier via the estimation of a piece-wise linear convex technology that envelops the data, as in Figure 1. The frontier technology encompasses the fully efficient units, while the efficiency of the rest of the units is measured as deviations from the frontier. A key drawback is that there is no random error; i.e., the approach is deterministic (although there have been attempts to introduce stochastic properties). Another weakness of this approach is its sensitivity to some factors:

- It is sensitive to outliers; i.e., units that are extreme in some sense (e.g. small offices run by workaholics)

- An increase in the number of outputs or inputs leads to an increase in efficiency scores. In small samples with many variables almost all units may be on the frontier.
- An increase in the number of production units yields a decrease in efficiency scores. Thus, a comparison of efficiency levels across industries requires adjustment for sample size; see Zhang and Bartels (1998).

It is not possible to determine which of the two major approaches dominates the other, since the true level of efficiency is unknown. Because in DEA all deviations from the frontier are interpreted as inefficiency, the SFA approach normally yields higher efficiency levels than DEA.

A more technical description of the two approaches is given in Appendix 1, which presents the mathematical formulation of the models applied in this report.

2.2 Productivity

The general definition of total factor productivity, TFP, is the ratio between an output index and an input index. This definition raises the question how outputs and inputs should be aggregated in the respective indexes. If a production function is used as an aggregator, the weights will be the marginal elasticities on both the output and the input side. This is the conventional aggregation approach.

Another type of aggregation is the proportional shrinking of an input vector and proportional expansion of an output vector. This is the aggregation performed in the so-called Malmquist productivity index discussed below.

We are often not so much interested in TFP per se, but in comparisons of TFP over time or across regions. Thus, TFP comparisons are binary comparisons between two observations in time or in space, while as we saw above, efficiency is measured against a common reference technology – the frontier.

For a parametric production function, the change in TFP over time is usually measured by the shift in production possibilities, which is the same as technical change. This is the time-trend approach. TFP is then the change in the output vector minus the change in the input vector; i.e., the residual factor of output growth net of changes in inputs. This is the definition used in the econometric, SFA, models discussed in this report.

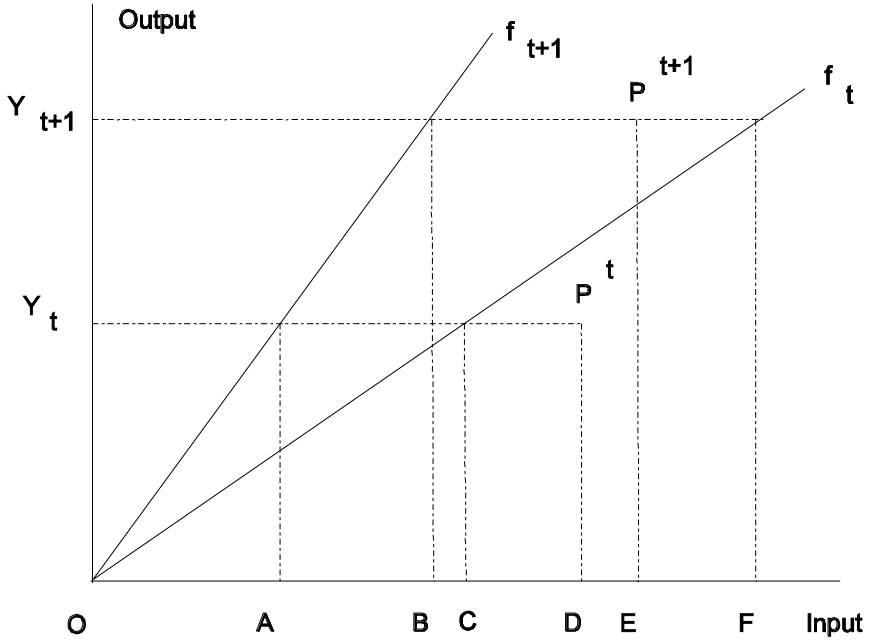
With DEA however, which is non-parametric, productivity change can not be investigated by a time trend approach. Instead, the Malmquist total productivity (MTP) index has come to dominate productivity research based on DEA frontiers.

The MTP index is only based on quantity variables, but nevertheless knowledge about at least one technology is required. This index and its parts can be expressed completely in terms of Farrell efficiency measures, so the need for technology information boils down to being able to calculate these measures. Thus, MTP fits very well as a complement to DEA[3].

Thus, the Malmquist productivity index offers an alternative way to measure productivity at a sectoral level. A graphical presentation of the total Malmquist productivity index is provided in Figure 2. Here we illustrate the construction of the input-oriented or input-based index.

[3] Originally, within a consumer theory context, Malmquist (1953) introduced the notion of *proportional scaling* of quantities observed in year t_2 to allow a consumer the same utility level as in year t_1. The proportional scaling factor was interpreted as a quantity index. Caves et al (1982) developed the Malmquist idea to a productivity index proper. However, the link to DEA was first developed by Färe et al (1994), (originally a working paper in 1989). They showed the relative ease with which the index could be calculated in the case of piecewise linear frontiers and since then there has been a rapidly increasing number of empirical applications. The additional attraction added by Färe et al (1994) was to maintain inefficient operations and decomposing productivity change into efficiency improvement and production frontier shift in the general case of multiple outputs and inputs, thus generalising the parametric single output approach in Nishimizu and Page (1982).

Figure 2. Construction of the Malmquist total productivity index, MTP



P is a production unit observed for two years, t and t+1. Between these time points the frontier function has shifted from f_t to f_{t+1}. In year t, technical efficiency for P^t measured against f_t is

$E^{t,t}$ = OC/OD

where the first index denotes the frontier year, and the second index denotes the observation. Measured against f_{t+1} technical efficiency for p^t is

$E^{t+1,t}$ = OA/OD

Correspondingly, technical efficiency for P^{t+1} relative to f_{t+1} is

$E^{t+1,t+1}$ = OB/OE

and measured against f_t it is

$E^{t,t+1}$ = OF/OE

The Malmquist input-based productivity index with frontier technology f_t as the reference may then be written as

$$M_t = \frac{E_{t,t+1}}{E_{t,t}}$$

The Malmquist input-based productivity index with frontier technology f_{t+1} as the reference may be written as

$$M_{t+1} = \frac{E_{t+1,t+1}}{E_{t+1,t}}$$

M >1 means productivity improvement, while M<1 means productivity decline.

The Malmquist index can be decomposed into two components, MC and MF_i.

MC is the change in efficiency relative to the frontier: i.e.,a catching up productivity index:

$$MC = \frac{\frac{OB}{OE}}{\frac{OC}{OD}} = \frac{E^{t+1.t+1}}{E^{t.t}}$$

MF is the change in distance to the frontier; i.e., a productivity index:

$$MF_i = \frac{E^{t,j}}{E^{t+1,j}}, \quad i,j = t,t+1 \quad i \neq j$$

and

A total productivity index, can then be defined as M_i = MC x MF_i:

For a fully efficient observation both years, MC = 1. The index is then a pure frontier – distance measure.

In general, the choice of reference technology is important for the empirical result. If there is no strong reason to choose a certain base, an alternative originally suggested in Färe et al (1994), is to take the geometric mean of M_t and M_{t+1},

$MG = [M_t\, M_{t+1}]^{1/2}$

In calculating the efficiency measures in Section 3.2 on which the subsequent Malmquist index is based, we have confined ourselves to the model with constant returns to scale. With regard to comparisons over time, this is fully plausible, since differences in returns-to-scale assumptions have greater implications for efficiency measures between production units in a cross section. However, in the other cases we apply both CRS and VRS models.

2.3 Interpretation of efficiency scores and productivity change

There are several reasons to be careful with interpretation of the empirical results of efficiency and productivity studies. For a start, what has come to be called efficiency measures is nothing but measures of distances from production units to a frontier. Especially in DEA with variable returns to scale, the frontier envelops closely around the data set. This means that a production unit may be at the frontier not because it is a well performing unit, but just because it lacks competition in that segment of the frontier. Typically, a closer look at a small frontier unit reveals it to be a rather poorly performing unit in terms of partial productivities. But given its size, it is still a frontier unit with efficiency score of 1. One should therefore check which units are so-called *self-evaluators,* in the sense that they lack competition, from other units. In this respect, constant returns to scale is more robust, and in general one should, expect rather large differences between efficiency scores generated by DEA-VRS compared to DEA-CRS for small and large units.

As discussed above, because of the inherent properties of the frontier techniques applied here, one should expect SFA to give higher mean efficiency, a smaller dispersion in a cross-section, and smaller variation over time than DEA. The reason for this is that, in the case of SFA, part of the variation in data is referred to random noise, while in DEA all deviations from the frontier are assumed to be inefficiency. Moreover, in the case of panel data estimation, another share of the variation in data is picked up the bank-specific effect. This component is meant to adjust for permanent differences in bank behaviour, which means that permanent inefficiency is caught by this component. Thus, the bank specific effect may also contain inefficiency, and then further reduce the explicit inefficiency scores.

In comparison with many other industries, one should also expect a lower degree of mean efficiency and a larger dispersion in efficiency

scores; i.e., between best practice and worst practice in the banking industry. The main reason for this is the rather low competitive pressure in the banking industry in combination with a mixture of banks with different types of ownership and objectives, commercial banks, cooperative banks, and not-for-profit savings banks. A related reason is the regulatory structure facing banks in different countries, which may further increase the distance between best practice and worst practice in international comparisons.

The rate of productivity change may also influence the distance between best practice and worst practice. During periods of rapid technical progress at the frontier, many production units may have difficult catching up and lag behind for some time, causing a decrease in mean efficiency. On the other hand, during periods of slow technical progress at the frontier, one would expect units below the frontier to catch up, causing an increase in mean efficiency.

Concerning interpretation of productivity change measures, one should bear in mind that, in general, productivity measures are rather model-dependent; see Kumbhakar et al (1999). This is also the case for productivity indexes. Usually Laspeyre and Paasche indexes constitute the outer bounds, with, for example the Tornkvist index in between. This is also true for the Malmquist index: The index based on the reference technology for year t (Laspeyre) often shows higher values than the index based on the reference technology for year t+1 (Paasche).

3 Empirical studies

3.1 Swedish branch office productivity change

3.1.1 The data

Analyses of productivity or efficiency at the branch office level are very rare; one of the few studies is Hjalmarsson et al (1991). This study covers the period 1983 to 1989 for the branch offices of one of the biggest Swedish commercial banks. The main reson for including this somewhat old study here is to show the large variation in productivity change between years at the bank office level and the large variation in efficiency among offices within a single bank. Data was derived from a computerised statistical system for labour-input measurement. The overall activity at the bank offices is specified and divided into about 700 points of measurement, then translated into time using the MTM standard. The various measurement points can be aggregated into a smaller number of activities or outputs. The following output variables are included in the analysis:

- *Deposits*; this refers to activities directed to the acquisition of savings and to related services.
- *Cashing;* this refers to cash-payment activity either in the bank offices or via a teller machine.
- *Loans*; this refers to activities related to the granting of credit.
- *Trusteeship* (Notariat); this refers only to the administration of assets which is related to the branch offices' depots. In the big cities, there are central notary (trust) departments that administer assets for a large number of customers. These departments are not included in the study.
- *Foreign Related;* this includes, foreign currency and traveller's cheques as well as payments to abroad undertaken directly by the office. The bulk of foreign transactions are executed at centrally

based and specialised departments using computers. These are left out of the analysis.

- *Fund;* this refers to that portion of the securities trade that is undertaken at the offices (delivery, swaps, payments etc.). Even here there exist centrally-placed functionaries who administer transactions and operate on the market (the stock exchange). These are not included either.
- *Lawyer*; this is a collective expenditure item used for those few offices which have their own bank lawyer.
- The following inputs are included:
- *Time worked*: full-time white-collar workers are assumed to put in 92400 minutes, in other words 1540 hours.
- *Information technology*, IT: costs for administrative data processing, measured in kSEK.
- *Office space:* measured in square meters.

Due to a statistical reorganisation between 1987 and 1988, these two years are not fully comparable. We thus divide the sample period into two parts: 1983-1987 and 1988-1989. The number of offices in the sample varies from 133 to 135.

We also note the following:

- 1984 and 1985 were years of low production across the board, the sharp decline having set in during 1983. This decline was reversed into an increase for 1986-87.
- Costs for information technology increased for all years.
- Office space, not unexpectedly, remained for the most part constant.

3.1.2 Empirical results

Average productivity growth

Table 1a shows productivity growth for a constructed average office (using the arithmetic mean), while mean (unweighted) productivity growth for the offices is presented in Table 1b.

Table 1a. Productivity changes for the average office, per cent

	MTP	MC	MF
1983/84	-22	-1	-21
1984/85	-6	1	-6
1985/86	34	0	34
1986/87	1	-4	5
1988/89	14	2	12

MTP is the Malmquist total productivity index, MC the catching-up – lagging-behind component and MF the frontier shift component derived in section 2.2. (Here, the first year is the reference year.)

Table 1b. Productivity changes for the office mean, per cent

	MTP	MC	MF
1983/84	-16	11	-24
1984/85	-8	-6	-1
1985/86	32	-4	37
1986/87	6	4	2
1988/89	13	4	8

The MTP value of -22 from 1983 to 1984, in Table 1a, indicates that total productivity fell by 22 per cent. The catching up component, MC, shows that the average fell slightly further behind the frontier, in spite of the large negative shift of the frontier itself, MF. After a notable decline in productivity between 1983-85, it rose by 34 per cent between 1985 and 1986, levelled out between 1986/87, and rose once again, by 14 per cent, between 1988 and 1989. Thus, the first two years show a negative productivity growth, which is hardly surprising given the sharp decline in all the output variables during these years. The MC measure suggests that the distance between the frontier and the average remained, roughly, constant over the sample period. The development of the unweighted arithmetic mean for the offices is slightly different from that of the constructed average office, reflecting the effects of the size distribution. The catching-up effect varies much more substantially here. The distance between the frontier and the average fell during 1983/84, rose during 1984-86, and fell for the remaining years of the sample. The differences between Tables 1a and 1b with regard to figures for 1983/84 indicate that it was the slightly larger offices on the frontier which experienced a reduction in efficiency.

Though there was a reduction in labour input from 1983 to 1985, computer input increased, while office-space remained constant. The decline in labour productivity was, therefore, much less than that for the other inputs; see Table 2, which presents results from a model where

labour is the only input, but where the outputs are the same as in the main model.

Table 2. Labour productivity change for the office mean, per cent

	MTP	MC	MF
1983/84	-11	15	-23
1984/85	-9	-4	-5
1985/86	29	-4	34
1986/87	7	0	7
1988/89	16	-1	16

The pattern is the same as that for total productivity although the labour productivity decline for 1983/84 was smaller. Moreover, labour productivity rose markedly in 1986/87, while MTP was almost constant.

Development of the individual offices

We have differentiated the individual offices with respect to their rate of productivity growth. Table 3 presents the number of offices with total productivity growth exceeding 5 per cent per year, the number from 5 to 2 per cent per year, those from 0 to 2 per cent per year, and the number of offices with negative productivity change.

The drastic change in productivity growth between the periods 1983/85 and 1985/89 is indisputable. While only a few offices had some productivity growth during 1983/85, the reverse is true for the second sub-period, though during 1986/87 almost half of the offices showed negative productivity growth.

Table 3. Number of offices in different ranges of productivity change

MTP-change/Year	1983/84	1984/85	1985/86	1986/87	1988/89
MTP > 5 %	2	14	123	40	89
5 % > MTP > 2 %	0	12	3	5	17
2 % > MTP > 0 %	1	8	3	11	6
MTP < 0 %	129	100	3	47	21

Efficiency

The spread of efficiency between offices together with the average office's efficiency; i.e., its distance from the best-practice frontier is highlighted in Table 4.

Table 4. Number of offices in different ranges of efficiency and structural efficiency each year

Efficiency/Year	1983	1984	1985	1986	1987	1988	1989
E = 1	42	30	51	36	29	36	29
1 < E > 0,90	23	29	23	26	18	20	34
0,90 < E > 0,80	36	43	32	33	39	32	34
0,80 < E > 0,70	25	24	26	28	22	33	28
0,70 < E > 0,60	5	6	10	8	24	10	7
E < 0,60	1	0	1	1	1	2	1
Structural efficiency	0.72	0.80	0.75	0.72	0.75	0.75	0.78

Structural efficiency during the period was within the interval 0.72-0.80. This implies that if all offices had been on the technology frontier observed output volumes could have been produced using between 20 and 28 per cent less inputs than was actually the case. Though this is a relatively common measure for structural efficiency in this type of analysis, the realisable productivity improvement potential is likely to be lower. The existence of a number of rigidities explains some of the differences in efficiency. Moreover, the offices are not as homogeneous as the analysis assumes.

The shape of the efficiency distribution shows relatively little variation. However, the number of frontier units varies considerably over the years.

3.1.3 Conclusions

The results are summarised as follows:

- Total factor productivity increased strongly during 1985-89
- Labour productivity grew somewhat more than total factor productivity
- Productivity growth was relatively uneven with a decline during 1983-1985, strong growth for 1985/86, low growth for 1986/87, and relatively strong growth for 1988/89.
- The distance between the frontier and the average has changed only minimally.
- Productivity growth varies markedly between branch offices.
- The structural efficiency measure suggests a productivity-improvement potential of 20-28 per cent; however, the realisable potential could be much lower.
- The spread of efficiency varies relatively little over the years.

3.2 Swedish banking efficiency and productivity change 1984-1995

In this section we report on a study, Battese et al (1997), investigating efficiency and productivity change in Swedish banks during 1984-95 using the SFA approach, with an input requirement frontier model, in which the amount of labour used is a function of the outputs and the quasi-fixed inputs, the number of branches and total inventories.

3.2.1 The data

The data set was obtained from the annual reports of Swedish banks. The data is not from a sample, but the entire population of Swedish banks. The data involved are for the years, 1984 to 1995. During the period some banks ceased to exist and others were established. Some of the changes were due to amalgamations of banks, or to take-overs.

The data is at the bank level rather than at individual branch level. Hence the large national banks with a large number of branches report only aggregate data. Other small regional banks may involve only one branch, and so the data are for that bank which is identical to a branch. Data for some small banks was deleted from the analysis because the values of one or more of the variables were missing or were clear outliers. This was particularly the case with some banks with only one or two part-time employees. A total of 1275 observations for 156 banks are used in the analyses reported below. In the first year, 1984, the number of banks in the data set was 140 and after a peak of 146 in 1985, the number declined to 65 in the last year, 1995. A total of 59 of the banks were observed in all twelve years.

The variables, *public loans*, *guarantees*, and *deposits*, are regarded as outputs. A key input variable of interest in a study of the efficiency of the banking industry is *labour*, because it is a significant component in the cost structure of banks. The efficiency of the use of labour in producing the relevant outputs, given the values of the quasi-fixed inputs, *branches* and *inventories*, is the focus of this study. Data on *Other Loans* and *Other Costs* were considered in preliminary analyses, but these variables were subsequently deleted because it was evident that there were problems with this data, perhaps due to changes in the definitions of the variables over the period. Data on variables, originally expressed in current money values, were deflated according to the values of the Consumer Price Index (CPI). Hence, all money values are expressed in terms of 1980 prices.

Labour use varied from about 800 to over 20 million employee hours per year. The values of public loans, guarantees, and deposits, also indicate the considerable variation in the size of the banks. The average number of years for which the banks were observed was about 10 years.

3.2.2 Empirical results

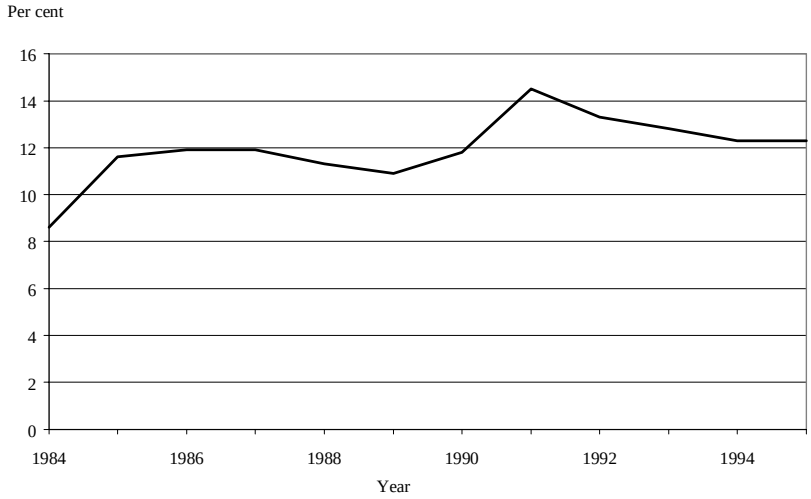
Efficiency

The individual technical ***inefficiencies*** of labour use for the banks range from 0 % to 166 % over the period. Thus, one bank was estimated to be fully technically efficient in labour use in at least one year, whereas another bank was estimated to use 166 % more labour than for a fully efficient bank with the same level of outputs and quasi-fixed inputs. The overall mean for the 156 banks was 12 % inefficiency. This is rather low, but one should keep in mind that a stochastic frontier model with bank-specific effects absorbs part of the inefficiency (in addition to the random noise component).

Of the 13 banks with inefficiency more than 20 % in the last year they were observed, 10 disappeared from the data set before 1995. Of the three surviving banks, two were savings banks and the other was a recently started international bank with a non-typical output structure.

Mean technical inefficiencies of labour use for the banks are graphed by year in Figure 3.

Figure 3. Mean technical inefficiency of labour use for Swedish banks, 1984-1995, per cent



Mean technical inefficiency tended to increase from 1984 to 1991, during the period of rapid expansion when the banks were in vigorous competition for market shares. After the crisis in 1992, mean technical inefficiency decreased slowly. Overall, there was a slight increase in mean technical inefficiency of labour use over the twelve-year period.

The results also indicate that the efficiency by the banks tend to:

- be larger for Föreningsbanken, Handelsbanken and Savings banks than for Other Banks;
- be smaller for SE-banken than for Other Banks;
- decrease as the number of branches increase;
- increase as the total inventories increase;

Because inventories are measured by book values, they reflect the modernity of these and especially for computer equipment with short depreciation periods. Thus, the efficiency-increasing effect of inventories should reflect the productivity-enhancing effect of more modern computer equipment.

The mean technical inefficiencies of the five types of banks, considered in the study, are given in Table 5. The difference between the two largest commercial banks is striking.

Table 5. Mean technical inefficiency of labour use by bank type, 1984-95

Type of Bank Mean	Technical Inefficiency (%)
1: Föreningsbanken (FB) (Co-operative; Particular bank	5.4
2: Handelsbanken (HB) (Commercial; Particular bank)	7.3
3: Sparbanker (SB) (Savings; Several banks)	11.3
4: SE-banken (SE) (Commercial; Particular bank)	30.4
5: Other Banks (Commercial; A few banks)	19.2

While Handelsbanken had an inefficiency level below 10 % for all years, SE-banken exceeded 20 % for all years, except 1984. During the banking crisis, Handelsbanken was never perceived as a crisis bank, while SE-banken was. One reason for the somewhat lower loan losses at Handelsbanken, compared with SE-banken and certain other Swedish banks, is usually attributed to the fact that Handelsbanken never abandoned its local branch networks as sources of decision-making about credit allocation. Moreover, the strong local authority at Handelsbanken branch-offices was combined with a system of profit sharing for bank staff.

Among the savings banks, there was substantial variation in inefficiency not reflected in the aggregated figure in Table 5. Most of these banks are local or regional banks operating as not-for-profit organisations. The same holds for the local or regional branches of Föreningsbanken. This bank specialised in credit to agriculture and small businesses before it merged with the largest savings bank in 1997.

Technical change and efficiency change

The rate of technical change (TC) in labour use is estimated to be negative for all years, but small in absolute values, see Figure 4. This indicates that there was modest technical *progress* in the *frontier* labour use during the years of our study. Starting at technical progress of 1.5 % in 1984, the industry continued to have technical progress, but at a decreasing rate, until 1992, when technical progress in labour use was effectively exhausted. Among the bank types, only Handelsbanken experienced technical *regress* in labour use over the entire period.

Inefficiency (IC) change of labour use over time was positive during 1984-1992, see Figure 4. This indicates that the inefficiency of labour use increased over those years. Thus the “average” bank did not manage to “catch up” with the labour-use frontier, which was experiencing technical progress. However, when technical progress in labour use was effectively exhausted in 1992, the banks began to “catch up”, i.e., the distance between observed labour use and the best-practice labour use

decreased somewhat. The dynamic interplay between changes in the frontier (technical change) and changes in inefficiency is discussed in more detail in Hjalmarsson (1973), Førsund and Hjalmarsson (1974) and Førsund, Hjalmarsson and Summa (1996).

One might have expected a much stronger impact of deregulation and the subsequent banking crisis on the efficiency of labour use in Swedish banks. An important reason why this did not occur might be the rapid return of Swedish banks to profitability which restored their access to capital markets. Most of the so-called "crisis banks" recovered rather rapidly; some were able to raise money from the capital market as early as 1993. Moreover, competitive pressure from abroad remained weak.

Figure 4. Technical change (TC) and inefficiency change (IC) in labour use by Swedish banks, 1984-1995

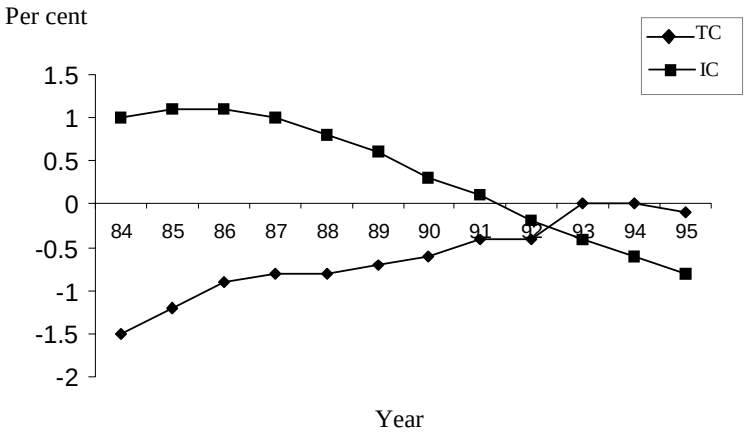
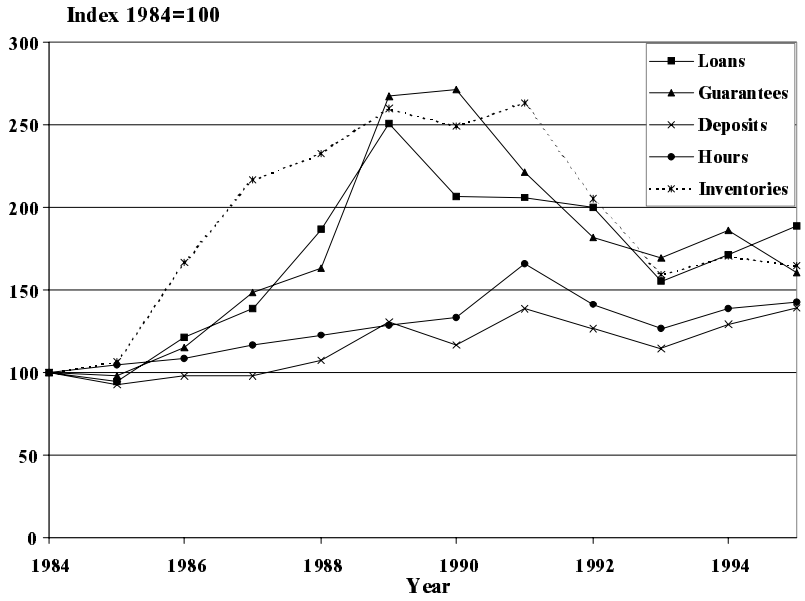


Figure 5 shows the relative movement of outputs, inventories and labour hours over the period, relative to 100 at the branch level in 1984.

Figure 5. Development of mean output and mean input



A major reason for the banking crisis in 1992 was the dramatic increase in the real rate of interest causing large changes in household and company portfolios with extensive repayment of old loans and less demand for new loans. Although deposits decreased somewhat in 1992, public loans, guarantees and inventories decreased substantially, and adjustments in the labour use was not at the same rate.

3.2.3 Conclusions

Inefficiency of labour use varied considerably for different banks over the period; overall inefficiency was estimated to be about 12 %. This implies that banks, on average, were using about 12 % more labour than if they were fully efficient, given the levels of outputs and quasi-fixed inputs.

An unexpected pattern emerges concerning the impact of deregulation and the subsequent banking crisis on technical change at the frontier. There was a decrease in technical change at the frontier and an increase in mean technical efficiency. While the development of efficiency seems to be driven by the frontier change, the change in the frontier is to a large extent driven by the development of output. Dramatic increases in the real rate of interest caused large changes in household and company portfolios, with extensive repayment of old, loans and less demand for

new loans. In general, the frontier banks did not manage to adjust the labour use in proportion to the decline in output.

3.3 The efficiency of Swedish banking in 1996

In addition to the analysis in the previous section, we have calculated DEA efficiency scores for Swedish banks in 1996. To investigate the sensitivity of the results with respect to the choice of variables, we compare the outcomes of different specifications under variable returns to scale (VRS) and constant returns to scale (CRS).

3.3.1 The data

The variables in this data set overlap with the previous one. We use the same basic three output variables, *loans, guarantees,* and *deposits,* but now we also add *net provisions,* i.e., the difference between commissions and other service charges received and the costs related to these. On the input side we use the *number of branch offices, inventories, other costs, credit losses, and labour costs, or alternatively, hours worked.* Thus, one important part of this section is to investigate the sensitivity of the results with respect to the choice of variables. Not including net provisions might "punish" especially the commercial banks or the larger banks. On the other hand, banks with large net provisions may also have more high-salary staff; therefore, we also use labour costs as a quality-adjusted labour input variable. See Table 6 for the differences between the models. We will also look at labour efficiency; i.e., we have calculated efficiency distributions for Models 1-4 but with only one input, labour.

Table 6. The difference between the models applied

Variable	Net provisions	Labour costs	Labour hours	Number of outputs	Number of inputs
Model 1	Yes	Yes		4	5
Model 2	Yes		Yes	4	5
Model 3		Yes		3	5
Model 4			Yes	3	5

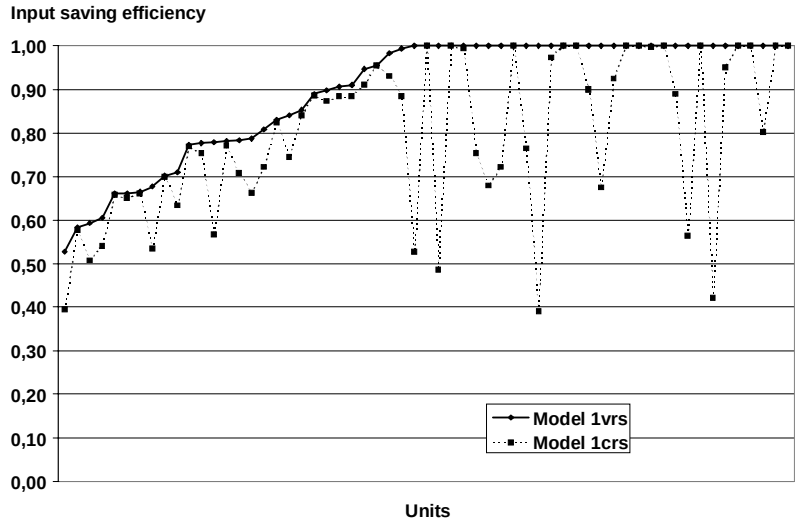
3.3.2 Empirical results

All inputs

Empirical results are illustrated in Figures 6-11. We regard Model 1 as the base model, while Models 2-4 are used for sensitivity analysis.

In general, with so many variables one would expect a rather large number of fully-efficient units (efficiency score = 1) under VRS, and this is also the case: Somewhat more than half of the units are on the frontier in Model 1; see Figure 6. On the other hand, under CRS, less than one fourth of the units are on the frontier. Mean efficiency decreases from 0.90 under VRS to 0.79 under CRS. All 17 units, which are on the frontier under VRS but not under CRS, are savings banks.

Figure 6. Input saving efficiency for Swedish banks, 1996: A comparison between variable and constant returns to scale



Reduction in the number of variables also has a strong impact on the number of frontier units; see Figure 8, where the number of outputs is decreased from 4 to 3, or Figures 10 and 11, with labour as the only input. In the former case the number of frontier units decreased from 31 to 25.

The choice between labour costs and labour hours has only a minor impact on the efficiency distribution; see Figures 7 and 9, in which Models 1 and 2 almost coincide, as do Models 3 and 4. Under VRS, four to six units (all savings banks) which are on the frontier in Model 1 become inefficient in the other models. Under CRS, four to seven banks become inefficient when moving from Model 1 to the other models; all

these are savings banks except Nordbanken which becomes inefficient in Models 2 and 4 (efficiency scores 0.94 and 0.88, respectively), i.e., when labour hours are substituted for labour costs.

On the other hand, there is a substantial difference between Models 1 and 2 on the one hand and Models 3 and 4 on the other; i.e., the results are very sensitive to the inclusion or exclusion of net provisions as an output variable. This is even more clear from Figure 8, which illustrates the difference between Model 1 and Model 3. Mean efficiency decreases from 0.89 in Model 1 to 0.73 in Model 3. All commercial banks (except Östgöta Enskilda Bank) are on the frontier in both models. For three units (Finn, Alingsås and Roslagen savings banks) there is a deviation in efficiency scores exceeding 0.50.

Figure 7. Input saving efficiencyfor Swedish banks, 1996, variable returns to scale

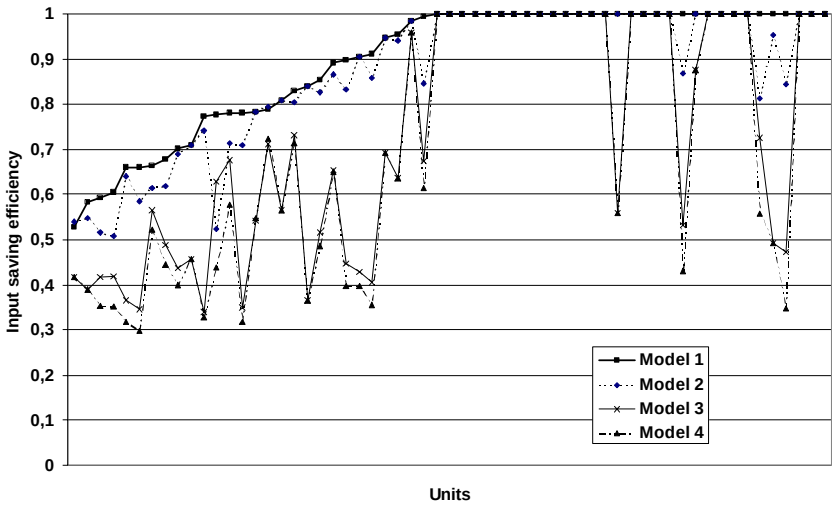
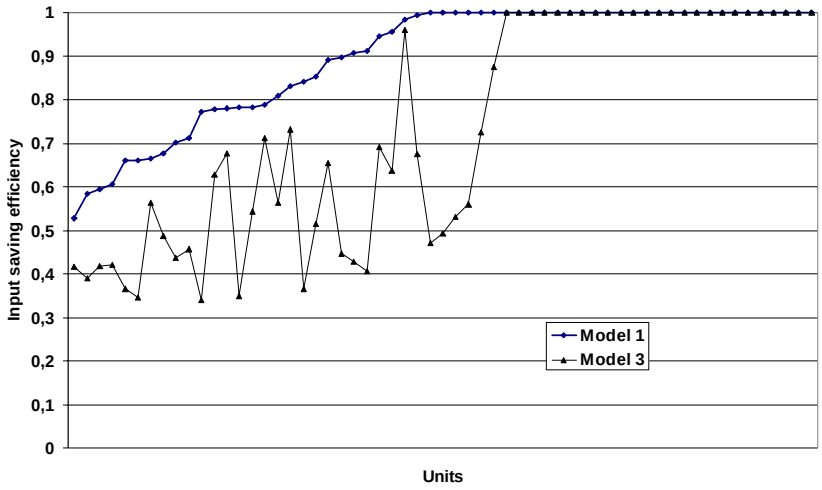


Figure 8. Efficiency distribution with and without net provisions as an output variable, variable returns to scale



Labour use efficiency

When labour is used as the only input, the number of frontier units decreases from 31 to 10 in Model 1 under VRS and from 14 to 3 under CRS; see Figures 10 and 11. Variation between the models is very large.Under VRS mean efficiency decreases from 0.79 in Model 1 to 0.45 in Model 4, and under CRS from 0.67 to 0.27. While all the commercial banks (except Östgöta Enskilda Bank) are on the frontier in Model 1 under VRS, only Handelsbanken and JP Bank are on the frontier in all four models. In CRS only Handelsbanken and JP Bank (along with Alingsås sparbank in Model 1 only) are on the frontier in all models, while SE-banken and Östgöta Enskilda Bank get efficiency scores of 0.75 and 0.62, respectively, in Model 1.

Figure 9. Efficiency distribution for Swedish banks, 1996. Constant returns to scale

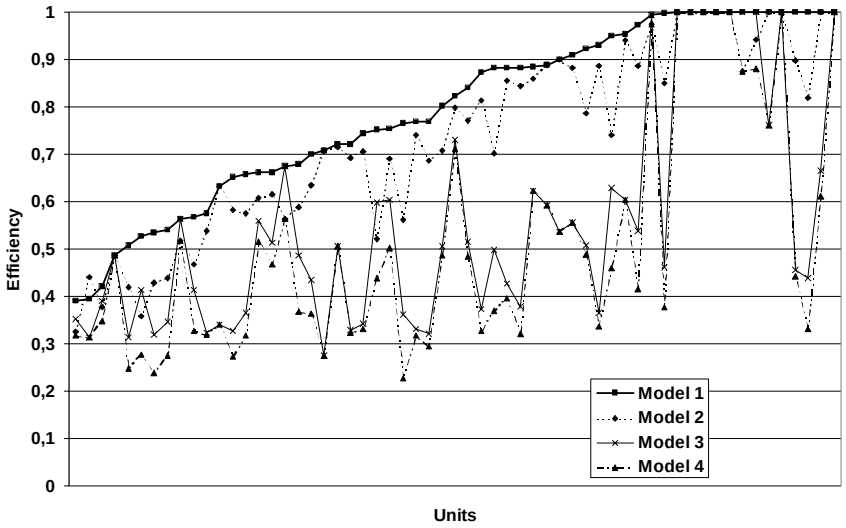


Figure 10. Labour use efficiency for Swedish banks, 1996. Variable returns to scale

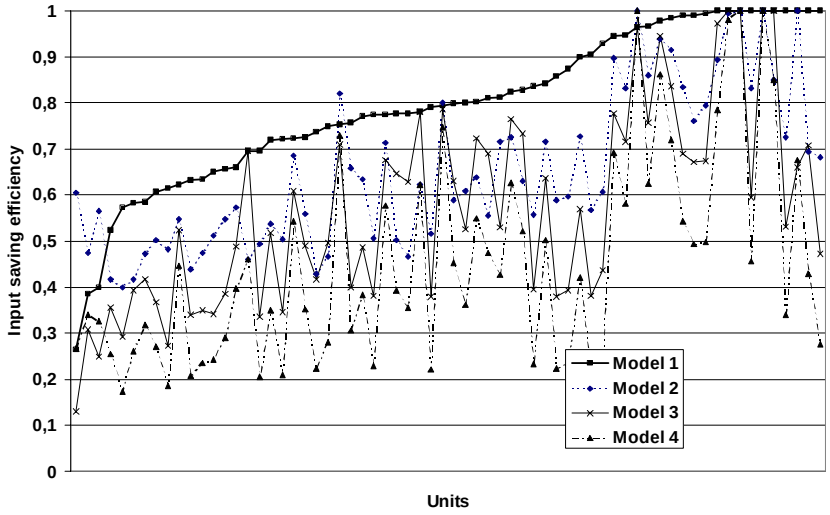
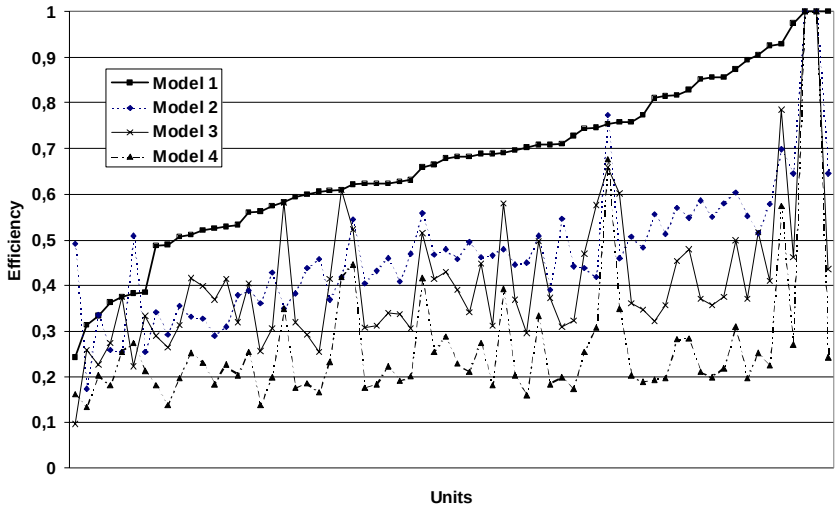


Figure 11. Labour use efficiency for Swedish banks, 1996. Constant returns to scale



3.3.3 Conclusions

Efficiency analysis of Swedish banks in 1996 illustrates the sensitivity of the results with respect to choice of variables. On the one hand, the choice between labour costs and labour hours has only a minor impact on the efficiency distribution, but on the other hand, the results are very sensitive to the inclusion or exclusion of net provisions as an output variable. While the results are robust for the commercial banks – which are all (with one exception) fully efficient – the differences between models for some savings banks are quite striking.

3.4 Banking productivity in a Scandinavian comparison

Because of low cultural barriers to entry into each other's markets, the performance of Swedish banking in comparison with its Scandinavian neighbours is of great interest. This issue was analysed in a few studies about 1990, and although there are plans to update these, this has not yet been realised; see Berg et al (1993) and Bukh et al (1995) which will be dealt with here.

3.4.1 The data

The data set in Berg et al (1993) consists of observations from 503 Finnish, 150 Norwegian, and 126 Swedish banks. Only a few, quite small, banks in each country have been omitted, due to data collection problems. The data were collected from official bank statistics in Norway and Finland, and from annual reports in Sweden, for the year 1990.

The output variables are *deposits loans,* and the *number of branches.*
The input variables are *number of hours worked* and the *book value of machinery and equipment.*

To convert monetary values into a common unit, both the official exchange rate and the purchasing- power parity rate as computed by the OECD were used, but it did not matter very much which rate was used.

In Bukh et al (1995) the data set for 1990 was extended to include 182 Danish banks and the variables were extended to include *guarantees* as an output, and *non-labour and non-capital operating expenses* as input.

3.4.2 Empirical results

Berg et al (1993) uses the Malmquist productivity index; Sweden is chosen as the technology base for productivity comparisons, due to the superior performance of its largest banks. A common technology base facilitates intercountry comparisons, but the choice of base country will influence the results. Productivity comparisons are shown in Table 7.

Table 7. Intercountry productivity comparisons, 1990, average and largest units in each country

	Malmquist productivity index (MTP)		Efficiency component (MC)		Technology component (MF)	
Scale	VRS	CRS	VRS	CRS	VRS	CRS
Norway/Finland:						
Average	12 %	9 %	7 %	-17 %	5 %	31 %
Largest	-18 %	19 %	0 %	-19 %	-18 %	47 %
Sweden/Finland:						
Average	63 %	52 %	48 %	39 %	10 %	9 %
Largest	35 %	51 %	0 %	43 %	35 %	5 %
Sweden/Norway:						
Average	46 %	40 %	38 %	68 %	5 %	-17 %
Largest	64 %	27 %	0 %	77 %	64 %	-28 %

The average Norwegian bank has about 10 % higher productivity than the Finnish average bank, 12 % under VRS, 9 % under CRS. For the largest bank the scale assumption is crucial, with 18 % higher Finnish productivity under VRS, as opposed to 19 % higher Norwegian productivity under CRS.

The Swedish average bank is much more productive than both the Finnish and the Norwegian average bank, 63 % higher than Finnish under VRS, 52 % under CRS, and 40 % and 46 % respectively, higher than Norwegian. The Swedish largest banks are also much more productive.

As discussed earlier the Malmquist productivity index can be decomposed into an efficiency component and a frontier technology component, the latter measuring the distance of the frontier technologies relative to the Swedish common-reference technology.

The large gap between Swedish and Finnish average productivity is mainly due to the efficiency component, although Sweden also has a technology advantage. The efficiency component explains even more of the Swedish-Norwegian gap. In fact, under CRS, Norway even has a technology advantage. Since the largest banks are on the frontier under VRS, there is in that case no efficiency component; i.e., its value is zero. There is just a difference in frontier technology levels.

The extension of the data set to include Danish banks and inclusion of one more output and input variable does not change the previous picture very much; see Table 8.

Table 8. Intercountry productivity comparisons, 1990. Malmquist productivity index, average and largest units in each country

	VRS		CRS	
	Average unit	Largest unit	Average unit	Largest unit
Norway/Finland	4 %	16 %	-6 %	2 %
Denmark/Finland	51 %	16 %	32 %	67 %
Denmark/Norway	46 %	0 %	40 %	63 %
Sweden/Finland	64 %	16 %	55 %	39 %
Sweden/Norway	59 %	0 %	64 %	36 %
Sweden/Denmark	9 %	0 %	18 %	-17 %

The average Swedish bank still has a higher productivity than the average bank in the other Scandinavian countries, now including Denmark, both under VRS and CRS.

3.4.3 Conclusions

In an inter-Scandinavian comparison, the Swedish banking sector in 1990 was much more productive than the Finnish, Norwegian, and Danish banking sectors. This conclusion also holds for the largest banks, except in comparison with Denmark. The large gap between Swedish and Finnish average productivity is mainly due to the efficiency component, although Sweden also had a technology advantage, which was the source of the Swedish-Danish productivity gap. The Swedish-Norwegian productivity gap is more model- dependent, and consists of both components.

3.5 The efficiency of Swedish banks in a European comparison

As the final step in this review of empirical studies, we draw on a study by Andersson (1999) on the relative efficiency of a set of large European banks observed in 1996.

3.5.1 The data

This analysis is based on a sample of 104 of the largest banks in eight countries, i.e., the Netherlands, Denmark, Belgium, France, Italy, Great

Britain, Sweden and Germany. The German, British, Italian and French banking markets constitute the largest markets in Europe, while the remaining four are representing the smaller market places; see Table 9. The complete list of banks is reported in Appendix 2.

Table 9. The distribution of banks

Country	Number of banks
Germany	21
Italy	17
UK	16
France	14
Sweden	11
Belgium	11
Denmark	9
Netherlands	5

Most of the countries in the study have a few core banks, which are dominant players in their domestic markets, but their importance is largely balanced by equally important regional banking groups, particularly in Germany, France, and Italy: Table 10 shows two size indicators for the sample.

Table 10. The size of the average bank in the sample, 1996

	1996	
Country	Assets MECU	Number of employees
Belgium	53 936	6 337
Denmark	22 881	2 781
France	127 002	25 431
Germany	120 575	12 299
Italy	48 151	13 018
The Netherlands	11 342	25 528
Sweden	22 842	3 240
UK	81 790	28 344

The main bank- types represented are commercial banks (67), savings/co-operative banks (26), and specialised banks (investment banks, mortgage banks-, and non-credit institutions) (11).

The average bank in the sample had some 14 000 employees and a branch network of almost 700 branches.

3.5.2 Empirical results

Here we report efficiency scores under both VRS and CRS, with two different choices of input variables and with the three output variables, *loans, number of branch offices*, and *deposits*. L-vrs and L-crs are the pure labour efficiency models i.e. hours worked as the only input under VRS and CRS respectively. X12-vrs and X12-crs include *inventories* as the second input under VRS and CRS respectively.

Because the number of variables is small in this data set, one would expect rather low levels of efficiency and a large dispersion between best practice and average efficiency. This is in fact the case. There is a large variation in efficiency.

The results are shown in Figures 12-15. The differences in mean efficiency scores are quite large. In the base model, X12-vrs mean efficiency varies from 0.31 for Italy to 0.70 for the Netherlands. Sweden ranks second from bottom with a mean efficiency score of 0.35. Sweden also ranks second from the bottom in the VRS labour-use model, with a mean efficiency score of 0.20 compared with 0.61 for the Netherlands.

Figure 12. Country mean efficiency scores for European banks, 1996

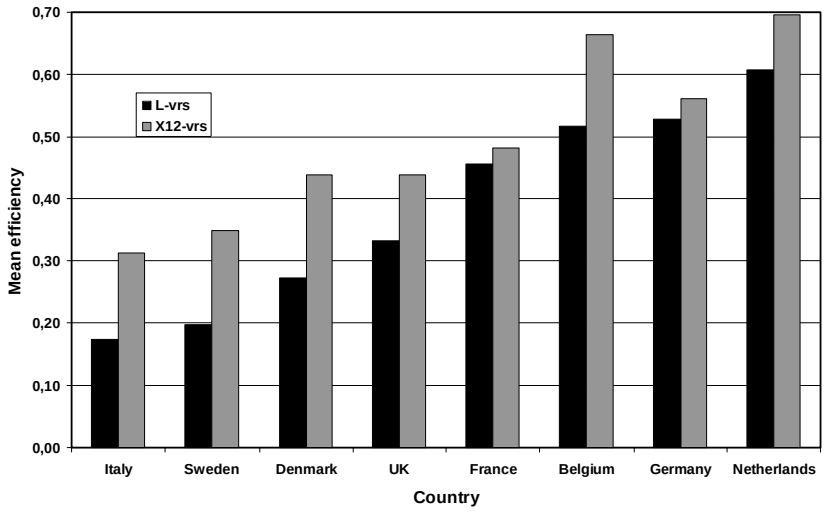
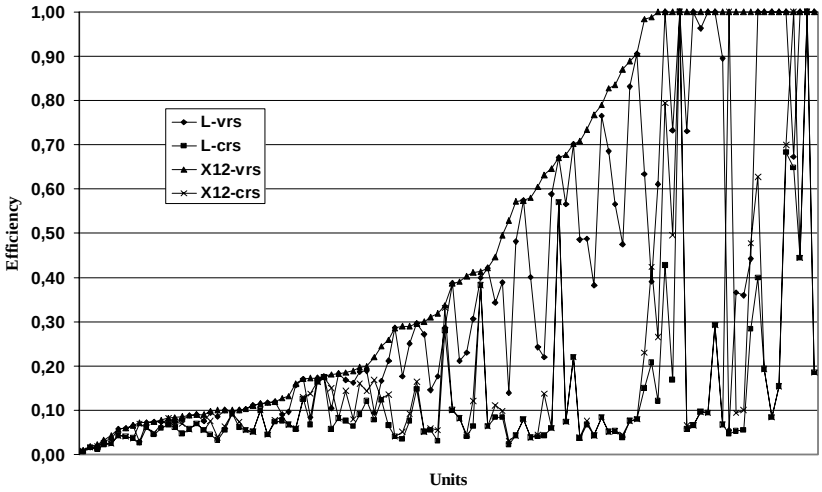


Figure 13 shows the efficiency distribution for the entire set of units.

Figure 13. The efficiency distribution of large European banks, 1996



The number of frontier units in the base model X12-vrs is 23, where as it is only 13 in the labour-use model under VRS. All countries are represented at the frontier, Sweden with only one bank.

The distance between best practice and worst practice is unusually large, with three worst performing UK banks at the very bottom (efficiency score 0.01-0.02) in all models. The worst-performing Swedish bank has an efficiency score of 0.07 both in the base model and the labour use model with VRS.

The location of the Swedish banks in the base model is shown in Figure 14 and in the labour use model in Figure 15. The Swedish banks are spread along the entire efficiency distribution in the two-input model. In the labour use model none of the Swedish banks has an efficiency score exceeding 0.50.

Figure 14. The location of Swedish banks in the efficiency distribution of European Banks, 1996, variable returns to scale

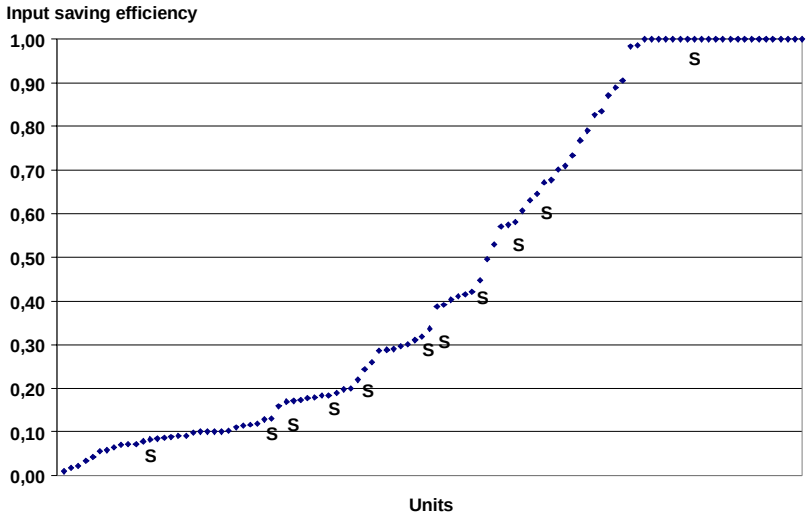
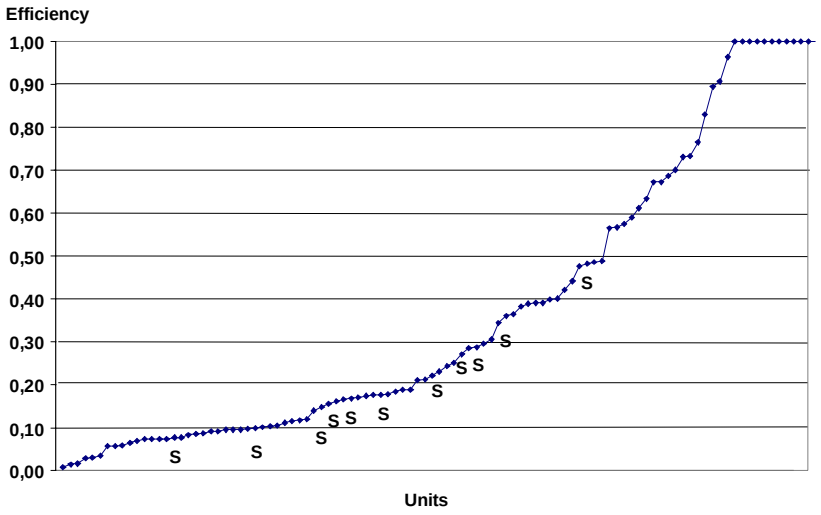


Figure 15. The location of Swedish banks in the labour-use efficiency distribution of European banks, 1996, variable returns to scale



3.5.3 Conclusions

There is a large variation in efficiency between banks, both within each country analysed and across countries. In a European perspective,

Swedish commercial banking efficiency on average ranks second from the bottom, both in a two-input model and in a labour-use model. Important caveats here are the relatively small sample of banks from each country included, and the small number of inputs and outputs covered by the study.

It is difficult to know the impact on the efficiency distribution of including more variables, on the input side as well as the output side. Since we have no additional data for a sensitivity analysis, one should be cautious in interpreting the results.

References

Aigner, D., J. and S.F. Chu, (1968), ”On Estimating the Industry Production Function”, *American Economic Review*, 58, 226-239.

Aigner, D., C.A.K. Lovell and P. Schmidt, (1977), ”Formulation and Estimation of Stochastic Frontier Production Function Models”, *Journal of Econometrics*, 6, 21-37.

Andersson I., (1999), ”Efficiency in European Banks”, Department of Economics, Göteborg University, Working paper.

Battese, G.E., (1997), ”A Note on the Estimation of Cobb-Douglas Production Functions when Some Explanatory Variables Have Zero Values”, Journal *of Agricultural Economics*, 48, 250-252.

Battese, G.E., and T.J. Coelli, (1995), ”A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data”, *Empirical Economics*, 20, 325-332.

Battese, G.E., Heshmati, A. and L. Hjalmarsson, (1997), ”Efficiency of Labour Use in the Swedish Banking Industry: A Stochastic Frontier Approach”, Paper presented at The Fifth European Workshop on Efficiency and Productivity Analysis, Copenhagen, October 9-11.

Berg, S.A., F.R. Førsund, L. Hjalmarsson and M. Suominen, (1993), ”Banking Efficiency in the Nordic Countries”, *Journal of Banking and Finance*, vol. 17 Nos. 2-3: 371-388.

Berger, A. N. and D. Humphrey, (1992), ”Measurement and Efficiency Issues in Commercial Banking”, In: Z. Grilliches eds., ”Output Measurement in the Service Sectors”, University *of Chicago Press: 245-279*, *Chicago*.

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

Bukh, P.N.D., S.A. Berg and F.R. Førsund, (1995), ”Banking Efficiency in the Nordic Countries: A Four-Country Malmquist Index Analysis”, Norges Bank, Arbeidsnotat.

Caudill, S.B., and J.M. Ford, (1993), ”Biases in Frontier Estimation and Firm-Specific Inefficiency Measures in the Presence of Heteroscedasticity”, *Economics Letters*, vol. 41, 17-20.

Caves, D., L. Christensen, and E. Diewert, (1982), ”The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity”, *Econometrica*, 50 (6), 1393-1414.

Charnes, A., W.W Cooper and E. Rhodes, (1978), ”Measuring The Efficiency of Decision Making Units”, *European Journal of Operational Research*, vol. 2, No. 2, 429-444.

Coelli, T.J., (1996), ”A Guide to FRONTIER Version 4.1: A Computer Program for Frontier Production Function Estimation”, CEPA Working Paper 96/07, Department of Econometrics, University of New England, Armidale.

Colwell, R.J, and E.P. Davis, (1992), ”Output and Productivity in Banking”, *Scandinavian Journal of Economics*, (Supplement), 111-129.

Farrell, M., J., (1957), ”The Measurement of Productive Efficiency”, *Journal of the Royal Statistical Society*: Series A, 120 (III), 253-281.

Fixler, D.J. and K.D . Zieschang, (1992), ”User Costs, Shadow Prices, and the Real Output of Banks” In: Zvi Grilliches eds., ”Output Measurement in the Service Sectors”, *University of Chicago Press*, 219-243.

Färe, R., S. Grosskopf, B. Lindgren, and P. Roos, (1994), ”Productivity Developments in Swedish Hospitals: A Malmquist Output Index Approach”, in A. Charnes, W.W. Cooper, A.Y. Lewin, and L.M. Selford (eds.), *Data Envelopment Analysis: Theory, Methodology and Applications*, Boston, Kluwer Academic Publishers.

Førsund, F.R., and L. Hjalmarsson, (1974), ”On the Measurement of Productive Efficiency”, *Swedish Journal of Economics* 76, 141-154.

Førsund, F.R., and L. Hjalmarsson, (1979a), ”Frontier Production Functions and Technical Progress: A study of General Milk Processing in Swedish Dairy Plants”, *Econometrica*, vol. 47, (4), 883-900, 1979.

Førsund, F.R., and L. Hjalmarsson, (1979b), ”Generalised Farrell Measures of Efficiency: An Application to Milk Processing in Swedish Dairy Plants”, *Economical Journal*, vol. 89, 294-315.

Førsund, F.R., L. Hjalmarsson and T. Summa, (1996), ”The Interplay between the Micro Frontier and Short-Run Industry Production Functions”, *Scandinavian Journal of Economics*, 98, 365-386.

Greene W.H., (1993), ”The Econometric Approach to Efficiency Analysis”, in *The Measurement of Productive Efficiency; Techniques and Applications, eds.*, Fried, H., in *eds.*,Fried, H., *The Measurement of Productive Efficiency: Techniques and Applications.*

Grosskopf, S., (1996), ”Statistical Inference and Nonparametric Efficiency: A Selective Survey”, *Journal of Productivity Analysis*, vol. 7, 161-176.

Heshmati, A., (1997), ”Labor Demand, Labor-Use, Efficiency and Risk in Swedish Saving-Banks, Department of Economics”, Göteborg University, Working Paper 1997.243.

Hjalmarsson, L., (1973), ”Optimal Structural Change and Related Concepts”, *Swedish Journal of Economics*, 75, 176-192.

Hjalmarsson, L., Bjurek, H., and A. Isaksson, (1991), ”Produktivitets-utveckling och effektivitet inom svensk bankverksamhet”, In *Produktivitet i vissa tjänstesektorer*, Expertrapport nr 2 till Produktivitetsdelegationen, Allmänna Förlaget.

Hoover, A.D., (1998), ”Swedish Savings Banks: A Sector Analysis of Labor Cost Efficiency Using Stochastic Frontier Analysis with Panel Data”, Department of Economics, Göteborg University Working paper, 1998.

Humphrey D., (1984), ”The U.S. Payment System, Costs Pricing, Competition and Risk”, Monograph Series in Finance and Economics, Solomon Bros Centre for Study of Financial Industry, NY.

Kumbhakar S., Heshmati, A. and L. Hjalmarsson, (1999), ”Parametric Approaches to Productivity Measurement: A Comparison Among Alternative Models”, *Scandinavian Journal of Economics,*101(3), 405-424.

Kumbhakar S., Heshmati, A. and L. Hjalmarsson, (1998), ”How Fast Do Banks Adjust? A Dynamic Model of Labour-Use with an Application to Swedish Banks”, Department of Economics, Göteborg University Working paper, 1998.

Lovell, C.A.K., (1993), ”Production Frontiers and Productive Efficiency”, in *The Measurement of Productive Efficiency; Techniques and Applications, eds*., Fried, H., in *eds*., Fried, H., *The Measurement of Productive Efficiency: Techniques and Applications.*

Malmquist, S., (1953), ”Index Numbers and Indifference Surfaces”, *Trabajos de Estatistica*, 4, 209-242.

Meeusen, W. and J. van den Broeck,(1977), ”Efficiency Estimation from Cobb-Douglas Production Functions With Composed Error”, *International Economic Review*, 18, 435-444.

Mester L. J., (*1997*), ”Measuring efficiency at U.S. Banks: Accounting for heterogeneity is important”, *European Journal of Operational Research*, 98 (2), 230-242.

Nishimizu, M and Page, J., M., (1982), ”Total factor productivity growth, technological progress and technical efficiency change: Dimensions of productivity change in Yugoslavia 1965-78”, *Economic Journal* 92, 920-936.

Richmond, J., (1974), ”Estimating the Efficiency of Production”, *International Economic Review*, 15, 515-21.

Sherman, J.D. and F. Gold, (1985), ”Bank Branch Operating Efficiency: Evaluation with Data Envelopment Analysis”, *Journal of Banking and Finance* 9. 297-316.

Triplett , J.E., (1992), ”Comments on "Measurements and Efficiency in Banking”, by A.N. Berger and D. Humphrey In: Zvi Grilliches eds., ”Output measurement in the service sectors”, *University of Chicago Press. 287-296*.

Tulkens, H., (1993), ”On FDH Efficiency Analysis: Some Methodological Issues and Applications to Retail Banking, Courts and Urban Transit”, *Journal of Productivity Analysis*, 4: 183-210.

Vassiloglou, M. and D. Giokas, (1990), ”A study of the Relative Efficiency of Bank Branches: An application of data envelopment analysis”, *Journal of operational research society*, 41, 591-97.

Wykoff, F. C., (1992), ”Comments on measurements and efficiency in banking”, In: Z. Grilliches ed., ”Output measurement in the service sectors”, *University of Chicago Press*. 279-287, Chicago.

Zhang, Y., and R. Bartels, (1998), ”The effect of sample size on the mean efficiency in DEA with an application to electricity distribution in Australia, Sweden and New Zealand”, *Journal of Productivity Analysis*, 9 (3), 187-204.

Appendix 1

The stochastic frontier model

We assume that the following translog stochastic frontier model defines labour use for the 156 banks in our data

$$\ell n\, Y_{it} = \beta_0 + \beta_0^{*} D_{it} + \sum_{j=1}^{6} \beta_j x_{jit} + \sum_{j \le k=1}^{6} \sum^{6} \beta_{jk} x_{jit} x_{kit} + V_{it} + U_{it} \qquad (1)$$

where the subscripts i and t represent the i-th bank (i = 1, 2, …, 156) and the t-th year of observation (t = 1, 2, …, 12);

Y represents the total quantity of labour used (in hours per year);

D is a dummy variable for guarantees, which has value one if guarantees were zero or not observed, and, otherwise zero [4]

x_1 is the logarithm of the total amount of public loans (in SEK1,000);

x_2 is the logarithm of the *maximum* of the guarantees (in SEK1,000);

x_3 is the logarithm of deposits (in SEK1,000);

x_4 is the logarithm of the number of branches;

x_5 is the logarithm of the value of machinery and equipment;

x_6 is the year of observation, where x_6 = 1, 2, …, 12;

the V_{it}s are random variables, associated with measurement errors in the labour variable or the effects of unspecified explanatory variables in the model, which are assumed to be independent and identically distributed with $N(0, \sigma_v^2)$-distribution, independent of the U_{it}s;

The U_{it}s are non-negative random variables, associated with inefficiency of labour use, given the levels of the outputs and the quasi-fixed inputs, such that U_{it} is obtained by the truncation (at zero) of the $N(\mu_{it}, \sigma^2)$-distribution; where

[4] Some small banks had zero guarantees. This variable permits the intercepts to be different for banks with positive and zero guarantees, see Battese (1997).

$$\mu_{it} = \delta_0 + \sum_{j=1}^{4} \delta_{0j} D_{jit} + \sum_{j=1}^{4} \delta_j z_{jit} \qquad (2)$$

where D_1, D_2, D_3, and D_4 are dummy variables which have value one if the observation involved is on a bank of types 1, 2, 3, and 4, respectively, and, otherwise zero;

z_1 is the logarithm of the number of branches (i.e., $z_1 = x_4$);

z_2 is the logarithm of total machinery and equipment (i.e., $z_2 = x_5$);

z_3 is the year of observation (i.e., $z_3 = x_6$); and

z_4 is the square of the year of observation (i.e., $z_4 = z_3^2$).

The model incorporates non-neutral technical change in the use of labour in the production of the outputs, given the quasi-fixed inputs in the banking operations.

The specification of the model for the inefficiency effects in equation (2) implies, in addition to accounting for the existence of inefficiency effects in the use of labour in producing outputs that the variations in the inefficiency effects are a function of other variables. In this case, we permit the level of the inefficiency effects in labour use to be different for five different types of banks (associated with the four dummy variables for types of banks), and they also depend on the number of branches in the banks and the year of observation.

The parameters of the stochastic frontier model, defined by equations (1) and (2), were simultaneously estimated by the method of maximum likelihood using the computer program, FRONTIER 4.1, written by Coelli (1996).

The technical efficiency of labour use for the i-th bank in the t-th year of observation (TE_{it}, given the values of the outputs and quasi-fixed inputs, is defined as the ratio of stochastic frontier labour use to observed labour use. *Stochastic frontier labour use* is defined by the value of labour use if the technical inefficiency effect, U_{it}, were zero (i.e., it the bank were fully efficient in the use of labour). Given the specifications of the translog stochastic frontier labour-use function in equation (1), it can be shown that

$$TE_{it} = \exp(-U_{it}) \qquad (3)$$

which indicates that technical efficiency is no greater than one. The reciprocal of this quantity, exp(U_{it}), is no less than one, and can be interpreted as a measure of the technical inefficiency of labour use.[5]

Computation of efficiency scores by DEA

The input-saving measure under VRS is found by solving the following LP problem for each plant, k, with output y_k and input x_k

$$\min_k E_{1k}$$

subject to the following restrictions: (4)

$$y_{rk} \leq \sum_j^N \lambda_{kj} y_{rj}, \quad r = 1,...,m \tag{4a}$$

$$E_{1k} x_{ik} \geq \sum_j^N \lambda_{kj} x_{ij}, \quad i = 1,...,n$$

$$\sum_j^N \lambda_{kj} = 1 \tag{4b}$$

$$\lambda_{kj} \geq 0, \quad j = 1,...,N \tag{4c}$$

where E_{1k} is the input-saving efficiency measure for plant k, m is the number of outputs, n is the number of inputs and N the number of plants.

Restriction (4a) implies that the reference plant must produce at least as much as plant k, while restriction (4b) implies that the efficiency-adjusted volume of input used by plant k must at least amount to the input volume used by the reference plant. Restriction (4c) is the condition for VRS. If this restriction is omitted, CRS is implied; E_1 and E_2 then coincide with each other.

[5] More precisely, the amount by which exp(U_{it}) *exceeds one* is a measure of technical inefficiency of labour-use, because it gives the proportion by which actual labour use exceeds the corresponding stochastic frontier labour value, given the level of outputs and quasi-fixed inputs.

Appendix 2

The selection of large European banks

SWEDEN

Stadshypotek
Wasabanken
Sparbanken
Föreningsbanken
Tryggbanken
Skandiabanken
JP Bank
Östgöta Banken
Nordbanken
Handelsbanken
SE-banken

GERMANY

Depfa Bank
Deutsche Bank
Bayerische Hypotheken und Wechsel Bank
Landesbank Schleswig Holstein
Stadtsparkasse Köln
Hamburger Sparkasse
Dresdner Bank
DG Bank
Commerzbank
Bayerische Stadtsparkasse
BHF Bank
Hamburgerische Landesbank
West LB
SGZ Bank
Bankgesellschaft Berlin
DSL Bank
Deutsche Ausgleichbank
Vereinsbank
Nord LB

Deutsche Girocentrale
Frankfurter Sparkasse

ITALY

Cassa di Risparmio di Verona
Banca Commerciale Italiana
Banca Ambrosiano Veneto
Cassa diRisparmio di Parma et Piacenza
Banca Populare di Novara
Banca Nazionale di Lavoro
Rolo Banca
Cassa di Risparmio
Credito Italiano
Banca di Roma
Banca Nazionale di Agricultura
Medio Banca
Banca Agricola di Mantova
Istituto Bancarioi di San Paolo
Istituto Mobiliare Italiana
Cassa di Risparmio di Firenze
Banca Monte Paschi di Siena

GREAT BRITAIN

Flemings Bank
Schröders
Lloyds Bank
HSBC
Moscow Norodny Bank
Standard Chartered Bank
The Royal Bank of Scotland
Nationwide
Bradford and Bingley
Abbey National
Hambros
Northern Rock
Nikko Bank
Barclays Bank
Bank of Scotland
National Westminster Bank

THE NETHERLANDS

ING
ABN
De Nationale Investeringsbank N.V.
Bank Nederlandske Gemeenten
SNS

DENMARK

Arbejdernes Landesbank
Sydbank
Bikuben
Realkredit
Jyske Bank
BRF Kredit
Den Danske Bank
Unidanmark
Nykredit Bank

BELGIUM

CGER-Bank
CERA Bank
ANHYP
IPPA Bank
Generale Bank
BBL
Kredietbank
Bank van Roeselare
Bacob Bank
Crédit à L= Industrie
DEXIA

FRANCE

Banques Populaires
Credit Locale de France
SOVAC
Credit Lyonnais
Credit Mutuel
Société Générale
Caisse de Depot Group

Banque Parisbas
Crédit Immobilier
Banque Worms
Banque Nationale de Paris
Banque Indosuez
Crédit Agricole
Crédit Commerciale de France