The Equilibrium Size of the Financial Sector[*]

Thomas Philippon[†]
New York University

August 2007

Abstract

Over the past 60 years, the value added of the U.S. financial sector has grown from 2.3% to 7.7% of GDP. I present a model of the equilibrium size of this industry and I study the factors that might explain its evolution. According to the model, a shift in the joint distribution of cash flows and investment opportunities across U.S. firms has increased the demand for financial services. Improvements in the relative efficiency of the finance industry also play a role. Without these improvements, a much larger fraction of firms would be financially constrained today.

Figure 1 shows the evolution of the share of the U.S. GDP accounted for by the Finance and Insurance industry over the post-war period. This share has grown from 2.3% to 7.7%. The share of the financial sector in the compensation of all employees has followed a similar trend. Why is the US devoting nearly 8% of its human capital to the provision of financial services? Understanding this evolution is the main motivation of this paper. To do so, I present a simple general equilibrium model of a production economy and I study the factors that pin down the supply and demand of financial services.

There exists a large literature on economic growth and structural change, and the rise in the finance industry is in some ways similar to the rise in other skill-intensive services.[1] There are, however, at least two reasons to pay close attention to the finance industry. The first reason is that this industry, unlike the rest of the service sector, would not exist in an Arrow-Debreu economy. If markets were complete, the health care industry would remain, but the finance industry would be trivial. How then should we interpret the dramatic growth of this industry over the past 60 years? Does this imply that the U.S. economy has drifted away from the Arrow-Debreu benchmark?

The finance industry is special for a second reason. The literature on structural change has emphasized non-homethetic preferences and the trade-off between home production and market production. These explanations may not be entirely relevant for financial services. In addition, it is far from obvious that differences in total factor productivity (TFP) growth between the manufacturing sector and the finance industry could account for the trend in Figure 1. The finance industry has benefited greatly from improvements in information technologies.[2] Moreover, a plausible benchmark is that it takes a given number of financiers to monitor a given number of entrepreneurs. If this is the case, the value added of financial services depends directly on the productivity of entrepreneurs, and technological progress in the industrial sector leads to balanced growth with a constant size for the finance industry.

The growth of the finance industry therefore raises particular issues, and requires a specific model. There are two approaches to modelling financial services. One approach assumes the presence of transaction costs and studies the organization of the industry. Financial institutions (FIs) are to financial products what retailers are to goods and services.

---

[1] Stigler (1956), Kuznets (1957), and Baumol (1967). See Buera and Kaboski (2006) for a recent contribution.

[2] Philippon and Resheff (2007) document a large reduction in the number of bank clerks since 1980.

However, as Freixas and Rochet (1997) argue, "the progress experienced recently in telecommunications and computers implies that FIs would be bound to disappear if another, more fundamental, form of transaction costs were not present". A second approach, which I follow in this paper, focuses instead on information asymmetries and moral hazard.[3]

I use a standard model of financing constraints, augmented with endogenous monitoring and career choices. The model economy has overlapping generations of agents and two sectors: industrial and financial. In the first period of their lives, agents decide in which sector they wish to work. The industrial sector produces goods that can be consumed or invested. All agents in the industrial sector engage in production, albeit with different levels of productivity. Some of them also receive an investment opportunity. Borrowing is limited by moral hazard, and some valuable investments may not be financed. Agents in the financial sector have access to a monitoring technology that reduces moral hazard. They sell their services to agents who would not be able to invest otherwise. In the model, the cost of financial intermediation is an opportunity cost: An agent cannot be a financier and an engineer at the same time.

A theory of the equilibrium size of the financial sector can shed light on a number of relevant issues. First, what determines the equilibrium demand for financial services? In particular, what is the role of the joint distribution of cash flows and growth options? Second, on the supply side, what are the consequences of productivity gains in the financial sector? Do they lead to a larger or a smaller finance industry? How do they affect the level of credit rationing? When productivity improves, does the financial sector grow until all credit constraints have disappeared? Third, what are the links between various measures of financial development? It is common in the empirical literature to measure financial development using private credit over GDP. Is this a good measure of financial development? How does it relate to the size of the financial sector measured with value added? Does an increase in the size of the financial sector signal that more firms are constrained, or fewer? To the best of my knowledge, no existing model addresses these issues in a simple unified framework.

This paper is also related to the literature on financial development and economic growth: Greenwood and Jovanovic (1990), Bencivenga and Smith (1991), King and Levine

---

[3]See Diamond (1984), Boyd and Prescott (1986) and Holmström and Tirole (1997) for instance.

3

(1993), and more recently Greenwood, Sanchez, and Wang (2007). My contribution, relative to these papers, is to study the size of the financial sector, and to propose an explanation for its evolution over time.

Section 1 presents the model and discusses in more details how it relates to the existing literature. Section 2 characterizes the equilibrium. Section 3 studies the determinants the equilibrium level of financial services provided in the economy. Section 4 presents a calibration and proposes an interpretation for the growth of U.S. financial sector since the 1950s. Section 5 concludes.

1 The model

1.1 Technology and preferences

Consider an economy with overlapping generations of agents who live for two periods. They work when they are young and discount the future at rate ρ. The size of each generation is normalized to 1. An agent $i \in [0,1]$ born at time t consumes $C_{1,t}^i$ when she is young and $C_{2,t+1}^i$ when she is old. The agent chooses a career and a consumption path in order to maximize her expected utility:

$$U_t^i = E_t \left[C_{1,t}^i + \frac{C_{2,t+1}^i}{1+\rho} \right]. \tag{1}$$

The economy has two sectors, industrial and financial. The industrial sector produces a good that can be consumed or invested. The financial sector produces monitoring services that are used by entrepreneurs of the industrial sector.

Career Choice

In their first period, agents choose a career. Let n_t be the number of agents who choose the industrial sector. The remaining $1 - n_t$ enter the financial sector. I start by describing an agent's career within the industrial sector. After she enters the industrial sector, an agent receives two endowment shocks: $\tilde{\theta} \in \{0, \theta\}$ is a binary random variable and $\alpha \in (0, \infty)$ is continuously distributed across agents. Both shocks are publicly observable. The first shock measures the investment opportunity of the agent. Investment requires x_t units of consumption good at time t and delivers $\tilde{\theta} x_t$ units of capital at time $t + 1$. The required scale x_t is common to all the projects started at time t. Each unit of capital produced at

4

time $t+1$ delivers one unit of consumption good per period, and depreciates at rate δ. Let π be the fraction of agents in the industrial sector who receive an investment opportunity:

$$\pi \equiv \Pr\left(\tilde{\theta} = \theta\right).$$

The second shock determines the productivity of the agent in the first period of her life. An agent who receives a shock α produces αx_t units of output. The exogenous scale/productivity parameter x_t generates a balanced growth path for the economy, as explained in section 2. Let $\bar{\alpha}$ be the unconditional mean of α:

$$\bar{\alpha} \equiv E\left[\alpha\right].$$

The parameter $\bar{\alpha} > 0$ measures the average productivity of the industrial sector, relative to its physical capital requirements. The shocks α and $\tilde{\theta}$ are correlated. Let $F^{\theta}\left(.\right)$ be the cumulative distribution of α *conditional on* $\tilde{\theta} = \theta$, and let $f^{\theta}\left(.\right)$ be the corresponding density function. The distribution of α conditional on $\tilde{\theta} = 0$ plays no role in the analysis and needs not be specified explicitly.

Production and capital accumulation

The consumption good is produced by agents in the industrial sector, with average productivity $\bar{\alpha} x_t$, and by capital, with a yield of one unit of consumption good per unit of capital. Let K_t be the stock of capital at the beginning of period t. The total amount of goods produced at time t is:

$$Y_t = \bar{\alpha} x_t n_t + K_t. \tag{2}$$

Capital accumulates over time according to:

$$K_{t+1} = \left(1 - \delta\right) K_t + \theta x_t e_t,$$

where e_t is the number of agents who invest. Since only a fraction π of agents in the industrial sector have a positive innovation opportunity, e_t cannot be more than πn_t.

Enforcement constraint and monitoring

Entrepreneurs have access to a stealing technology. After investing x_t unit at time t, an entrepreneur can always default at time $t+1$ and steal a fraction z of her project, while

5

the remaining $\theta x_t - z x_t$ is lost. The parameter $z \in [0, \theta)$ captures the severity of the moral hazard problem. Moral hazard may prevent the entrepreneur from obtaining the necessary funds for the initial investment. Monitoring services can be used to alleviate the financing constraint. If m units of monitoring services are allocated to an entrepreneur, the fraction she can steal is reduced to $z - m$. The monitoring services are supplied by the finance industry, with a productivity of μ units of services per agent employed. The total amount of monitoring available in the economy at time t is therefore:

$$\mu \left(1 - n_t \right).$$

1.2 Discussion of the model and the literature

The model has several special features, which were chosen to make the analysis transparent and tractable. The first main assumption concerns the role of the financial sector. Levine (2005) defines five broad functions of the financial sector: (i) to produce information and allocate capital ex-ante, (ii) to monitor investments and exert corporate governance, (iii) to facilitate trading and diversification, (iv) to mobilize and pool savings and (v) to ease the exchange of goods and services. In this paper, I focus on the monitoring of corporate investment and the allocation of capital. The demand for financial services is entirely due to moral hazard, and I abstract from transaction costs and trading frictions. For the purposes of this paper, there is not much difference between (i) and (ii). The task of the financial sector is to make sure that lenders are repaid. Whether this is achieved by the ex-ante selection of good projects, or by the ex-post enforcement of contracts and good governance, does not really matter for the analysis in this paper.

One might question the extent to which the services of the financial sector reflect project selection, monitoring, enforcement or governance. This seems like a reasonable description of the activities of investment banks, venture capitalists and private equity funds, and the part of commercial banking that is not related to consumer lending and residential mortgages. On the other hand, the model probably does not capture well the activities of passive mutual funds. In terms of value added, however, these mutual funds represent only a small fraction of the financial sector. If it were only for these funds, the approximation of an Arrow-Debreu economy would be quite accurate. An open question is how to model the role of hedge funds. There is evidence that hedge funds intervene directly in the governance

of companies (Brav, Jiang, Partnoy, and Thomas (2007)). They can also improve the allocation of capital by arbitraging mispricing across stocks.

Modeling financial intermediaries as monitors has a long tradition in economics and finance, and much work has been done on the issue of who monitors the monitors (Diamond (1984), Holmström and Tirole (1997)). I abstract from this issue by not introducing any asymmetric information or moral hazard between savers and financial intermediaries. This implies that firm boundaries within the financial sector are irrelevant. All that matters is the productivity of the sector, measured by the parameter μ.

Regarding preferences, the main assumption is that agents are risk neutral. As explained in the introduction, this paper looks at the role of the financial sector in providing services to the corporate sector. The paper does not take into account the services provided to households, in terms of liquidity and personal insurance. Risk neutrality also implies that the real rate of interest is constant as long as consumption is interior for at least some agents.

On the production side of the model, there are two important assumptions. First, in equation (2), k and n are perfect substitutes. This assumption is made for convenience: It implies a fixed relative price, and it simplifies the analysis by reducing the dimension of the system.[4] Second, growth is driven by the exogenous process x_t: This is not a model of endogenous growth. As a consequence, financial development has only a level effect on output per capita, and not a growth effect. I chose this framework because my goal is to provide a benchmark analysis of the determinants of supply and demand for financial services. The model could easily be extended to allow for endogenous growth along the lines of Romer (1986) or Lucas (1988).

Finally, the model assumes a closed economy. This means that the entire demand for financial services comes from domestic firms. I return to this issue in interpreting the results of the calibration in section 4.

---

[4]One should be careful in interpreting n as labor and k as capital, however. In the model, α is the current income of agents in the industrial sector, but agents are not necessarily workers. They can also be interpreted as firms. Reciprocally, the income of entrepreneurs comes from their contribution to the future stock of k, but in the data, part of this income could be registered as labor income.

2 Equilibrium

Market clearing

Equilibrium in the goods market requires that consumption plus investment equals total production. Let $C_{1,t}$ be the total consumption of the young agents, and let $C_{2,t}$ be the total consumption of the old agents. The market clearing condition is:

$$C_{1,t} + C_{2,t} + xe_t = K_t + \bar{\alpha} x_t n_t. \tag{3}$$

In their second period, old agents receive dividends equal to K_t. They also own the stock of capital that remains after depreciation. Let q_t be the ex-dividend price of one unit of capital. Without bequest motives, the second period consumption of the old generation equals its total income:

$$C_{2,t} = K_t + (1 - \delta) q_t K_t. \tag{4}$$

Note that $C_{1,t}$ and $C_{2,t}$ are aggregate quantities and that agents within a generation typically have different levels of consumption. Combining (3) and (4), we obtain the saving equation:

$$\bar{\alpha} x_t n_t - C_{1,t} = (1 - \delta) q_t K_t + xe_t. \tag{5}$$

The left-hand-side of equation (5) measures the savings of the young generation. The right-hand-side of the equation measures the investments of the young generation, which consist of buying the existing machines from the old generation, and financing new projects.

Asset prices

Let r_t be the interest rate between period t and $t + 1$. The ex-dividend price of one unit of installed capital satisfies the dynamic equation:

$$q_t = \frac{1 + (1 - \delta) q_{t+1}}{1 + r_t}. \tag{6}$$

The net present value of a project per unit invested is:

$$v_t = \theta q_t - 1. \tag{7}$$

The total value is $v_t x_t$.

Balanced growth

I focus on the balanced growth path of the economy. Total factor productivity x_t grows at rate γ:

$$x_{t+1} = (1 + \gamma) x_t. \tag{8}$$

Let us use lower-case letters to denote the quantities scaled by productivity: $k_t \equiv K_t / x_t$, $c_t \equiv C_t / x_t$, and so on. Capital accumulation, written with scaled variables, becomes:

$$(1 + \gamma) k_{t+1} = (1 - \delta) k_t + \theta e_t$$

The scaled quantities are constant along the balanced growth path of the economy. In particular, the stock of capital is $k = \theta e / (\gamma + \delta)$. The interest rate is constant and the ex-dividend price of one unit of installed capital is simply: $q = 1 / (r + \delta)$. The saving equation (5) becomes:

$$c_1 = \bar{\alpha} n - g(r) e, \tag{9}$$

where the function $g(r)$ is defined by:

$$g(r) \equiv 1 + \frac{1 - \delta}{r + \delta} \frac{\theta}{\gamma + \delta}. \tag{10}$$

In equilibrium, the interest rate and the investment demand must adjust so that c_1 is always positive.

I now describe the steady states of three economies. First, I characterize the equilibrium of an economy without enforcement constraints. Second, I describe the equilibrium of an economy with moral hazard and without intermediation. The discussion of these two benchmarks is brief and is mainly used to provide the foundations for the analysis of the third economy, where moral hazard and active monitoring are both present.

2.1 No moral hazard

In this section, I consider the case where $z = 0$. Projects are funded if and only if they are profitable, that is, if v defined in equation (7) is positive. There are no financiers, and $n = 1$. The investment demand (in units of entrepreneurship) is therefore:

$$e^d(r) : \left\{ \begin{array}{ll} = \pi & \text{if } r + \delta < \theta \\ \in [0, \pi] & \text{if } r + \delta = \theta \\ = 0 & \text{if } r + \delta > \theta \end{array} \right\} \tag{11}$$

9

The saving curve, derived from equation (9) and the constraint that c_1 be positive, is:

$$e^s(r): \begin{cases} = 0 & \text{if } r < \rho \\ \in [0, \bar{\alpha}/g(\rho)] & \text{if } r = \rho \\ = \bar{\alpha}/g(r) & \text{if } r > \rho \end{cases} \tag{12}$$

The interest rate is pinned down by the market clearing condition: $e^s(r) = e^d(r)$. The following proposition characterizes the equilibrium without moral hazard, which is also depicted on Figure 2:

Proposition 1 *Without moral hazard, investment is strictly positive if and only if $\theta > \rho + \delta$. All projects are financed when $(1 + \gamma)\pi \leq (\delta + \gamma)\bar{\alpha}$. Otherwise $r = \theta - \delta$, entrepreneurs are indifferent between investing and not investing, and $e < \pi$.*

One important feature of this economy is that the equilibrium is independent of the conditional distribution of income $F^\theta(.)$. Only the unconditional mean $\bar{\alpha}$ matters. From now on, I assume that the no-moral hazard economy has a strictly positive investment rate with an interior solution for consumption. This holds under the following parameter restriction:

Assumption 1: $\theta > \rho + \delta$ and $\bar{\alpha} > \pi g(\rho)$

2.2 Moral hazard without monitoring

I now consider the case where $z > 0$, but I still assume that there is no active financial sector (one can think of this case as $\mu = 0$). Borrowing is limited by the fact that the entrepreneur can default. Consider an entrepreneur who has borrowed an amount b in her first period. Since each unit of capital (before depreciation) yields one unit of dividend each period, the cum-dividend value per unit is $(1 + r)/(r + \delta)$. If the entrepreneur defaults, she gets $zx(1 + r)/(r + \delta)$ in the second period of her life. If she does not default, she gets $(1 + r)\theta x/(r + \delta) - b(1 + r)$. The maximum amount of borrowing allowed in the first period is therefore $b^{\max} = (\theta - z)x/(r + \delta)$. An entrepreneur with current income αx can finance her investment if and only if $\alpha x + b^{\max} > x$. This defines a threshold α_h for financing without monitoring:

$$1 - \alpha_h \equiv \frac{\theta - z}{r + \delta}. \tag{13}$$

Internal cash must cover the difference between the pledgeable value of the project and the capital expenditures required to start the project. Agents with an investment option but whose income is less that α_h are financially constrained.

Let $e^c(r)$ be the effective investment demand curve under moral hazard. When $r + \delta > \theta$, it collapses to zero, just like in the case of no moral hazard examined in the previous section. When $r + \delta \leq \theta$, the constrained investment demand is given by:

$$e^c(r) = \pi \left(1 - F^\theta(\alpha_h)\right). \tag{14}$$

When $r + \delta = \theta$, the effective demand curve is vertical and e^c can be anywhere between 0 and $\pi \left(1 - F^\theta \left(\frac{z}{\theta}\right)\right)$. The saving equation is still given by (12), as in the case of no moral hazard. The equilibrium is depicted on Figure 3 and characterized by the following proposition:

Proposition 2 *Under assumption 1, with moral hazard and no monitoring, the interest rate is ρ and the number of project financed is $e^c(\rho)$.*

Proof. Under assumption 1, equation (9) shows that $c_1 > 0$. Therefore $r = \rho$. Investment is pinned down by the constrained demand schedule e^c evaluated at $r = \rho$. ∎

2.3 Active monitoring

I now turn to the case where firms have the option to hire monitoring services in order to relax their credit constraints. Suppose that an entrepreneur with current income α hires m units of monitoring. The pledgeable income becomes $x(\theta - z + m)/(r + \delta)$. The amount of monitoring required for this entrepreneur to be able to invest is:

$$m(\alpha) = (r + \delta)(\alpha_h - \alpha). \tag{15}$$

Let ϕx be the price of one unit of monitoring. For intermediation to be valuable to entrepreneurs, the price of financial services must be such that $v - m\phi \geq 0$. It is profitable for an entrepreneur to hire monitoring services if α is more than α_l, defined as the solution to:

$$m(\alpha_l)\phi \equiv v. \tag{16}$$

Let us now turn to the supply of financial services. In their first period, agents choose freely which sector they would like to work. Therefore, in any equilibrium where at least some

11

agents become financial intermediaries, the following indifference condition must hold:

$$\mu\phi = \bar{\alpha} + \pi\left(1 - F^\theta\left(\alpha_h\right)\right)v + \pi\int_{\alpha_l}^{\alpha_h}\left(v - \phi m\left(\alpha\right)\right)dF^\theta\left(\alpha\right). \tag{17}$$

The left-hand-side of equation (17) is the value of entering the financial sector. The right-hand-side is the value of entering the non-financial sector, which contains three terms. The first term is the expected income received when working in the industrial sector. The second term is the expected value of becoming an entrepreneur who can finance herself directly. The last term is the expected value of becoming an entrepreneur who hires financial services. Combining (16) and (17), I obtain:

$$\frac{\mu}{\pi} = \left(\frac{\bar{\alpha}}{\pi v} + 1 - F^\theta(\alpha_l)\right)m(\alpha_l) - \int_{\alpha_l}^{\alpha_h}m\left(\alpha\right)dF^\theta\left(\alpha\right). \tag{18}$$

Finally, equilibrium in the monitoring market requires that:

$$\mu\left(1 - n\right) = \pi n\int_{\alpha_l}^{\alpha_h}m\left(\alpha\right)dF^\theta\left(\alpha\right). \tag{19}$$

The difference with the previous case of moral hazard without monitoring is that the saving curve in Figure 3 also shifts as people move in and out of the industrial sector. Figure 4 describes the monitoring equilibrium.

Proposition 3 *Under assumption 1, the balanced growth path with active monitoring is characterized by $r = \rho$, and by α_l and n that solve equations (18) and (19). Credit rationing persists as long as $\alpha_l > 0$, and the size of the financial sector is strictly positive as long as $f^\theta\left(\alpha_h\right) > 0$.*

Proof. Under assumption 1, equation (9) shows that $c_1 > 0$. Therefore $r = \rho$. The RHS of equation (18) goes to zero as $\alpha_l \to \alpha_h$ so for any value of $\mu > 0$, it is possible to find $\alpha_l < \alpha_h$ that solves equation (18). If the density $f^\theta\left(\alpha_h\right) > 0$ and if f^θ is continuous, then the RHS of (19) is strictly positive and n is strictly less than one. QED. ∎

Even when μ is very small, some agents always choose to become financial intermediaries. Of course, this means that ϕ must become arbitrary large. But close to the cutoff α_h, the amount of monitoring required to obtain financing is arbitrary small, and some entrepreneurs are always willing to buy this small amount, even at a very high price. Therefore, the banking sector is active for any positive value of μ.

3 Comparative statics

I organize the comparative statics in the following way. I first study the parameters specific to moral hazard and financial intermediation, i.e. the parameters that affect the equilibrium of the economy with intermediation, but do not appear in the equilibrium of the economy without moral hazard. I then study the parameters that determine the macroeconomic investment opportunity set.

Three parameters affect the supply and demand of financial services in the model economy with moral hazard and intermediation, but would be irrelevant in an economy without moral hazard. On the supply side, μ measures the productivity of the financial sector. On the demand side, z measures the degree of moral hazard, while $F^\theta(.)$ measures the extent to which firms with growth options also have high current income.

Let us consider the supply side first. How does the size of the financial sector depend on μ? The following two propositions characterize the effects of μ. Consider first the extreme cases where the productivity of the financial sector is either very high or very low. Looking at equation (19), it seems difficult to predict the limiting behavior of the size of the financial sector, since the quantity of monitoring on the right hand side goes to zero when μ goes to zero. The behavior of n is therefore unclear. It turns out, however, that one can prove a general result, that holds for all distribution functions:

Proposition 4 *The size of the financial sector goes to zero when its productivity becomes either very small or very large.*

Proof. See appendix. ■

On the one hand, an increase in μ implies that the same amount of monitoring can be performed by fewer financiers. On the other hand, the drop in the price of monitoring services leads to a surge in demand. These two forces determine the effects of changes in the productivity of monitoring services. When n is close to one, the supply effect is negligible and the demand effect dominates. Therefore, starting from a value of μ close to zero, which implies a value of n close to one, as we have shown in the previous proposition, we see that an increase in μ leads to a decrease in n. For a very large value of μ, the density $f^\theta(\alpha_l)$ must eventually be close to zero. Therefore, n must increase in response to

an increase in μ. For intermediate values of μ, the comparative statics obviously depend on the shape of the density function f^θ. The demand effect dominates if the density at α_l is high enough, and an increase in μ then leads to an increase in the size of the financial sector. Finally, note that the increase in n might occur even though there is still a positive density at the cutoff α_l. Therefore the financial sector might shrink despite the fact that some firms remain rationed. Of course, one should keep in mind that the supply of monitoring services rises in any case, since α_l always decreases with μ.

Proposition 5 *Productivity gains in the financial sector always reduce credit rationing. Productivity gains lead to a larger financial sector when the productivity of the financial sector is relatively low. When the productivity of the financial sector is already high, further productivity gains decrease the size of the sector. In all cases, it is possible for the size of the financial sector to decrease in response to productivity gains even though some firms remain rationed.*

Proof. See appendix. ∎

Consider now a change in z. The change in z affects the self-financing cutoff α_h. This changes the monitoring function as well as the boundaries of integration. As expected, when moral hazard worsens, credit rationing increases. The impact on the size of the financial sector is ambiguous, however. On the one hand, it takes more resources to monitor a given set of firms. On the other hand, the pool of firms that are actually monitored shrinks.

The second parameter that determines the demand for financial services is the distribution function F^θ. Remember that F^θ is the distribution of α conditional on $\tilde{\theta} = \theta$. Thus, a change in F^θ while keeping the unconditional mean $\bar{\alpha}$ constant, is like changing the cross-sectional correlation between investment opportunities and current cash flows, while keeping aggregate productivity constant. The comparative statics for F^θ are unambiguous, unlike those for z. A decrease in f^θ, that is, a shift of F^θ to the right must always increase α_l. This is because the shadow value of an extra unit of income, conditional on having an investment opportunity, is more than one. Therefore, a right shift in F^θ, even holding $\bar{\alpha}$ constant, makes it more attractive to work in the industrial sector. To keep agents indifferent between the two sectors, the price of financial services must increase, and α_l must

14

increase. This reduces the demand for financial services in two ways. It shrinks the range of values of α for which monitoring is purchased, and it decreases the number of firms who need monitoring for a given monitoring region.

Proposition 6 *A decrease in the correlation between current income and investment options decreases α_h and increases the size of the financial sector. Worsening of moral hazard increases credit rationing but has an ambiguous effect on the size of the financial sector.*

Proof. See appendix. ∎

I now turn to the parameters that affect the investment opportunity set: π and θ. A shift in π is like a broadening of the investment set: more firms can invest, but the average productivity does not change. The increase in π increases the attractiveness of the industrial sector, therefore the price of financial services must rise to satisfy the indifference condition. This means that α_l must increase. Thus, rationing increases. The impact on n is ambiguous, since more firms would like to invest, but a smaller fraction can actually do so. A shift in θ is like a deepening of the investment set: each project becomes more valuable, but the number of projects stays constant. The impact on n is ambiguous, since more firms are monitored, but conditional on α, monitoring decreases.

Proposition 7 *An increase in θ, or a decrease in π, decreases rationing and has ambiguous effects on the size of the financial sector*

Proof. See appendix. ∎

4 A quantitative investigation

I focus my calibration on the U.S. economy because of the availability of firm level data over a long period. Greenwood, Sanchez, and Wang (2007), on the other hand, study a large cross-section of countries. These two approaches are complementary. The analysis of Greenwood, Sanchez, and Wang (2007) suggests that financial development can explain a significant fraction of the differences in income per capita between countries. I provide an explanation for the growth of the leading finance industry over time.

4.1 Empirical evidence and estimation of the parameters

The data used in the estimation are presented in Table 1. They cover the period 1956-2005, split into five sub-samples. I start by choosing standard values for the discount rate and for the depreciation rate:

Variable	Empirical Value	Model Parameter
Length of 1 Period	15 years	
Annual real rate	4%	$\rho = 1.04^{15} - 1$
Annual growth rate	2%	$g = 1.02^{15} - 1$
Annual depreciation rate	6%	$\delta = 1 - 0.94^{15}$

Next, I use the fact that the book value of a realized project is x while its market value is $\theta x / (\rho + \delta)$. I obtain θ by assuming a ratio of market value to book value of two.[5] It seems *a priori* difficult to calibrate the remaining parameters of the model, since they are not observable. Notice, however, that the equilibrium equations (18) and (19) depend only on the ratios μ/π and $\bar{\alpha}/\pi$. To further reduce the number of degrees of freedom in the calibration, I consider families of functions $f^\theta(.)$ that depend on only one parameter. I perform the calibration using either a uniform distribution or a downward sloping triangular distribution, both on the interval $\left[0, \alpha^\theta\right]$. The unknowns are therefore: z, μ/π, $\bar{\alpha}/\pi$ and α^θ. I estimate the benchmark parameters using data from the first decade, 1956-1965. I now explain how I use four observed quantities to pin down these parameters.

Capital Expenditures

In the model economy, (scaled) aggregate investment expenditures are equal to $\left(1 - F^\theta(\alpha_l)\right)\pi n$, and the stock of capital along the balanced growth path is:

$$k = \left(1 - F^\theta(\alpha_l)\right)\frac{\pi n \theta}{\delta + \gamma}.$$

The gross domestic product is $\bar{\alpha}n + k$. Therefore the investment share of GDP depends only on α_l and on the ratio $\bar{\alpha}/\pi$:

$$\frac{1 - F^\theta(\alpha_l)}{\bar{\alpha}/\pi + \left(1 - F^\theta(\alpha_l)\right)\theta/(\delta + \gamma)}$$

---

[5] This is a typical value for the US over the post-war period. Fama and French (2001) report an asset weighted average of 1.4 over the period 1963-1998. Hennessy, Levy, and Whited (2007) calibrate their model with an average value of 2.5 over 1968-2003, and they report values above 3 for firms that either issue equity or are likely to be credit constrained.

16

I calibrate this ratio to 11% using private non-residential fixed investment divided by private value added from the National Income and Product Accounts.

Investment share of low cash firms

Recall that x_t is the amount of capital required to start a project, while αx_t is the current income of the agent. Thus, α captures the ratio of income to capital expenditures, for the firms that actually invest. To make progress here, I need to use some information about the distribution of firms and investments. In the model, for tractability, I have assumed a unique fixed scale x_t for all the projects, but this is not a realistic assumption. I use a simple statistic to compare the model and the data: the share of total investment accounted for by firms whose income is less than one third of their capital expenditures. This statistic is useful because it does not involve taking a ratio of income over capital expenditures. It is robust to the discrepancy between the assumptions made in the model and the fact that firm sizes are very heterogenous in the real world. It also builds in the fact that large firms are more relevant than small firms. In the model, the investment share of low cash firms is $\left(F^\theta \left(0.33 \right) - F^\theta \left(\alpha_l \right) \right) / \left(1 - F^\theta \left(\alpha_l \right) \right)$.

I use Compustat to estimate this share over time. I use all firms in the industrial Compustat files with non missing values for income before extraordinary items and capital expenditures, and I exclude financial firms and firms in real estate. I compute:

$$s_t = \frac{\sum_i capex_{it} * (income_{it} < 0.33 * capex_{it})}{\sum_i capex_{it}}$$

To avoid issues with the timing of income and investment, in the above formula, $capex_{it}$ and $income_{it}$ are the sum of capital expenditures and income in year $t-2$, $t-1$ and t. The share of investment accounted for by firms with $\alpha < 0.33$ is shown on Table 1. It is has increased over time.[6] The cutoff must be substantially less than one because the relevant information is in the left tail of the distribution, but the particular value of 0.33 is arbitrary. To check the robustness of the calibration, I have performed the same exercise using 0.25 and 0.15 instead

---

[6] One might worry about a change in the coverage of Compustat. However, after 1975, the ratio of the number of employees covered by Compustat to total non-farm payrolls is constant. It is true that these firms are younger, but this is exactly what one would expect if financial markets improve. See Philippon and Sannikov (2007).

of 0.33, and obtained very similar results. The evolution of these three shares is shown on Figure 5. All three have increased over time. This new stylized fact is an important piece of information for the calibration. Firms that do not have enough income to cover their capital expenditures require financial services. Thus, Figure 5 suggests an increase in the amount of financial services over time. Notice, however, that this could be either because of shift in the underlying correlation between cash flows and investment options, captured by the function F^θ, or because of a shift in the cutoff α_l for a given distribution F^θ. In other words, the stylized fact could reflect either a supply shift, a demand shift, or a mixture of both. A quantitative model is needed to make sense of the data.

Corporate borrowing

In the model, firms with $\alpha < \alpha_l$ cannot invest, and firms with $\alpha > 1$ can finance their investment entirely with their current income, as depicted on Figure 4. Total corporate borrowing in each period is therefore equal to:

$$\pi n x \int_{\alpha_l}^{1} (1 - \alpha) f^\theta (\alpha) \, d\varepsilon$$

I compute the ratio of outstanding credit market instruments over GDP from the Flow of Funds Accounts of the U.S. This ratio is shown in Table 1. It has increased over time. The relevant variable for the calibration is the amount of new borrowing in each period. I assume that the average maturity of credit market instruments is 10 years, and I calibrate the model to a borrowing ratio equal to 1/10 of the outstanding value from Table 1. For robustness, I have redone the analysis assuming average maturities from 8 to 15 years.

Estimation

Finally, I match $1 - n$ as the value added share of the financial sector, also from Table 1. I estimate the baseline parameters using the average values over the first decade, 1956 to 1965. The estimated values, assuming that f^θ is uniform, are:

z/θ	μ/π	$\bar{\alpha}/\pi$	α^θ
0.84	4.43	5.34	1.21

The estimated values, assuming that f^θ is downward sloping triangular, are:

z/θ	μ/π	$\bar{\alpha}/\pi$	α^θ
0.82	3.42	4.90	1.80

18

4.2 Quantitative predictions

As explained above, the model is calibrated using values for the period 1956-1965. In this section, I investigate the quantitative properties of the model, in two ways. First, I compare the equilibrium for different values of the monitoring productivity parameter μ. Second, I create predicted values for the period 1966-2005 and I compare the predictions with the actual values.

Impact of productivity gains

Figure 6 shows the impact of changing the productivity of the financial sector. All the parameters of the model are the ones calibrated for the period 1956-1965 assuming a uniform density f^θ. The horizontal axis is normalized by the benchmark value for μ, called μ_0. The left panel shows the size of the financial sector, $1 - n$. The right panel shows the fraction of firms that are credit constrained. Consistent with proposition 5, Figure 6 shows that the size of the financial sector increases with μ when μ is small, and decreases with μ when μ is large. The right panel shows the fraction of firms that are credit constrained. The calibrated value is 12.87%. Productivity gains in the financial sector decrease credit rationing, as explained in proposition 5. While this is virtually undetectable, the size of the financial sector actually starts to decrease before the fraction of constrained firms reaches zero. This is qualitatively consistent with 5, but it appears quantitatively irrelevant. To a good approximation, productivity gains in finance should increase the size of the financial sector until all credit constraints are alleviated. Further gains are used to reallocate labor to the industrial sector while keeping enough financiers to make sure all positive NPV projects are financed.

Out-of-sample predictions

How well does the model predict the evolution of the finance industry from 1966 to 2005? I construct predicted values while keeping the technological parameters and the moral hazard parameter z constant. I let the conditional distribution of income, determined by α_t^θ, and the relative efficiency of the financial sector, μ_t, change over time. I pin down their values in the following way. First, I keep the investment share of GDP constant, which appears consistent with the evidence.[7] Second, I match the share of investment by low cash firms

---

[7]In the model, without changing any other technological parameter, this implies that the fraction of

from Table 1. I do not take into account the transitional dynamics and I assume that the economy is on its balanced growth path in each decade. Figure 7 shows the actual and predicted size of the financial sector. Figure 8 shows the actual and predicted ratio of credit market instruments to GDP. Note that the predictions do not use any direct information on the size of the finance industry or the corporate bond market beyond 1965. Up to the 1990s, the model seems able to predict most of the changes in the size of the financial sector. The model falls short in the more recent decade. Since the calibration forces all the demand for financial services to come from domestic firms, the discrepancy could reflect the globalization of the finance industry (Bekaert, Harvey, and Lumsdaine (2002)).

Counter-factual experiments

Suppose there had been no shift in productivity or in the correlation between growth options and cash flows. What would have happened? Table 2 shows the results of counter-factual experiments using the calibrated model. Keep in mind that the model is non-linear, so the effects are not additive: the total change is not the sum of the two partial counter-factual changes. Most of the increase in the size of the industry seems to come from a shift in the correlation between investment opportunities and current income. Even without productivity gains, the model predicts an increase in the size of the industry by 2.91 percentage points of GDP. Improvement in the financial sector only add another 0.54 percentage points to the size of the sector relative to GDP. However, without these improvement, the fraction of rationed firms would have increased substantially, from 12.87% to 19.23%.

5 Conclusion

This paper builds a model of the size of the financial sector. The model sheds light on the consequences of efficiency gains in the finance industry and on the determinants of the corporate demand for financial services. The equilibrium size of the financial sector depends on two main parameters: the productivity of the monitoring technology used by the financial sector, and the joint distribution of cash flows and investment opportunities across industrial firms. Because the efficiency of the financial sector affects the realized correlation between actual investment and actual income, as well as the size of the finance

constrained firms also stays constant.

industry, a quantitative model is needed in order to recover the structural parameters of the economy. According to the model, a shift in the distribution of investment opportunities towards low cash firms has increased the demand for financial services over the past fifty years. This demand shift accounts for a large part of the increase in the size of the finance industry. At the same time, improvements in the efficiency of the financial sector have prevented financial constraints from becoming more prevalent.

The existing literature has provided much evidence on the role of financial frictions by looking at the investment behavior of individual firms. It is fair to say, however, that there is not much consensus regarding the quantitative importance of these frictions for the whole economy, because it is hard to aggregate the various studies into one meaningful number. This paper offers a different perspective on the matter: If financial frictions were not important, why would we spend nearly 8% of GDP on financial services?

The analysis presented here is only a first step, however, mainly because it does not take into account the demand for financial services by households or the globalization of the finance industry. Incorporating these two forces is a task for future research.

References

BAUMOL, W. J. (1967): "Macroeconomics of Unbalanced Growth: The Anatomy of the Urban Crisis," *American Economic Review*, 57, 415–426.

BEKAERT, G., C. R. HARVEY, AND R. LUMSDAINE (2002): "Dating the Integration of World Capital Markets," *Journal of Financial Economics*, 65, 203–249.

BENCIVENGA, V. R., AND B. D. SMITH (1991): "Financial Intermediation and Endogenous Growth," *Review of Economic Studies*, 58, 195–209.

BOYD, J. H., AND E. C. PRESCOTT (1986): "Financial Intermediary Coalitions," *Journal of Economic Theory*, 38, 211–232.

BRAV, A., W. JIANG, F. PARTNOY, AND R. THOMAS (2007): "Hedge Fund Activism, Corporate Governance, and Firm Performance," *Journal of Finance*, forthcoming.

BUERA, F. J., AND J. P. KABOSKI (2006): "The Rise of the Service Economy," Working Paper, Northwestern University.

DIAMOND, D. W. (1984): "Financial Intermediation and Delegated Monitoring," *Review of Economic Studies*, 51, 393–414.

FAMA, E. F., AND K. R. FRENCH (2001): "Disappearing Dividends: Changing Firm Characteristics or Lower Propensity to Pay?," *Journal of Financial Economics*, 60, 3–43.

FREIXAS, X., AND J.-C. ROCHET (1997): *Microeconomics of Banking*. MIT Press, Cambridge.

GREENWOOD, J., AND B. JOVANOVIC (1990): "Financial Development, Growth, and the Distribution of Income," *Journal of Political Economy*, 98, 1076–1107.

GREENWOOD, J., J. M. SANCHEZ, AND C. WANG (2007): "Financing Development; The Role of Information Costs," Working Paper, University of Pennsylvania.

HENNESSY, C. A., A. LEVY, AND T. M. WHITED (2007): "Testing Q Theory with Financing Frictions," *Journal of Financial Economics*, 83, 691–717.

HOLMSTRÖM, B., AND J. TIROLE (1997): "Financial Intermediation, Loanable Funds and the Real Sector," *Quarterly Journal of Economics*, 112, 663–691.

KING, R. G., AND R. LEVINE (1993): "Finance and Growth: Schumpeter Might Be Right," *Quarterly Journal of Economics*, 108(3), 717–737.

KUZNETS, S. (1957): "Quantitative Aspects of the Economic Growth of Nations: Industrial Distribution of National Product and Labor Force," *Economic Development and Cultural Change*, 5, 1–111.

LEVINE, R. (2005): "Finance and Growth: Theory and Evidence," in *Handbook of Economic Growth*, ed. by P. Aghion, and S. N. Durlauf, vol. 1A, pp. 865–934. Elsevier, Amsterdam.

LUCAS, R. E. (1988): "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22, 3–42.

PHILIPPON, T., AND A. RESHEFF (2007): "Skill Biased Financial Development: Education, Wages and Occupations in the U.S. Finance Industry," Working Paper, New York University.

PHILIPPON, T., AND Y. SANNIKOV (2007): "Financial Development, IPOs, and Business Risk," mimeo, NYU.

ROMER, P. M. (1986): "Increasing Returns and Long Run Growth," *Journal of Political Economy*, 94, 1002–1037.

STIGLER, G. J. (1956): "Trends in Employment in the Service Industries," General Series. NBER.

Appendix

A   Proof of proposition 4

Consider first the limit when $\mu \to 0$. First, rewrite (17) as

$$\frac{\mu}{\pi m(\alpha_l)} = \frac{\bar{\alpha}}{\pi v} + 1 - F^\theta(\alpha_l) + \int_{\alpha_l}^{\alpha_h} \left(1 - \frac{m(\alpha)}{m(\alpha_l)}\right) dF^\theta(\alpha).$$

As μ goes to zero, $m(\alpha_l)$ also goes to zero and $\alpha_l \to \alpha_h$. We need to evaluate the limit of the integral. For all $\alpha \in [\alpha_l, \alpha_h]$, we know that $0 < m(\alpha) < m(\alpha_l)$ and therefore:

$$\left\| \int_{\alpha_l}^{\alpha_h} \left(1 - \frac{m(\alpha)}{m(\alpha_l)}\right) dF^\theta(\alpha) \right\| \le F^\theta(\alpha_h) - F^\theta(\alpha_l).$$

We can see that the integral goes to zero as μ goes to zero and:

$$\lim_{\mu \to 0} \frac{\mu}{\pi m(\alpha_l)} = \frac{\bar{\alpha}}{\pi v} + 1 - F^\theta(\alpha_h).$$

Using equation (19), we see that $(1 - n)/n = \pi/\mu \int_{\alpha_l}^{\alpha_h} m dF^\theta$. Since the monitoring function m is decreasing in $\bar{\alpha}$, it follows that

$$\left\| \int_{\alpha_l}^{\alpha_h} m(\alpha) dF^\theta(\alpha) \right\| \le m(\alpha_l) \left(F^\theta(\alpha_h) - F^\theta(\alpha_l)\right).$$

Since we have shown that $\mu/m(\alpha_l)$ has a finite limit, it follows that $1 - n \to 0$. In the other limit when $\mu \to \infty$, the result is clear from equation (19) since the integral of the right-hand-side is bounded by $\int_0^{\alpha_h} m(\alpha) dF^\theta(\alpha)$. QED.

B   Comparative statics

Let $\Delta[.]$ denote the total difference of a function or a variable of interest. To prove the various propositions, I differentiate equation (18):

$$\Delta\left[\frac{\mu}{\pi}\right] = \left(\Delta\left[\frac{\bar{\alpha}}{\pi v}\right] - \Delta\left[F^\theta\right](\alpha_l)\right) m(\alpha_l) + \left(\frac{\bar{\alpha}}{\pi v} + 1 - F^\theta(\alpha_l)\right) \Delta[m_l] - \int_{\alpha_l}^{\alpha_h} \Delta\left[m dF^\theta\right],$$

and equation (19):

$$\left(\int_{\alpha_l}^{\alpha_h} m dF^\theta + \frac{\mu}{\pi}\right) \frac{\Delta[n]}{n} = \frac{1-n}{n} \Delta\left[\frac{\mu}{\pi}\right] - \int_{\alpha_l}^{\alpha_h} \Delta\left[m dF^\theta\right] + m(\alpha_l) f^\theta(\alpha_l) \Delta[\alpha_l].$$

These formula hold because the boundary terms with α_l cancel out and because $m(\alpha_h) = 0$. The monitoring function from equation (15) can be written as:

$$m(\alpha) = z - \theta + (\rho + \delta)(1 - \alpha).$$

The total difference of this equation is:

$$\Delta[m] = \Delta[z] - \Delta[\theta] - (\rho + \delta)\Delta[\alpha] + \Delta[\rho + \delta](1 - \alpha).$$

B.1 Proof of proposition 5

Consider the impact of a change in μ. From equation (18), we see that

$$\Delta \left[\frac{\mu}{\pi} \right] = -\left(\rho + \delta \right) \left(\frac{\bar{\alpha}}{\pi v} + 1 - F^\theta(\alpha_l) \right) \Delta \left[\alpha_l \right]$$

So it is clear that α_h decreases with μ. The effect on the size of the finance industry, however, is ambiguous. From the monitoring market clearing (19), we see that

$$\left(\int_{\alpha_l}^{\alpha_h} m dF^\theta + \frac{\mu}{\pi} \right) \frac{\Delta \left[n \right]}{n} = \frac{1 - n}{n} \Delta \left[\frac{\mu}{\pi} \right] + m(\alpha_l) f^\theta(\alpha_l) \Delta \left[\alpha_l \right]$$

The sign of the RHS clearly depends on the value of n.

B.2 Proof of proposition 6

For a change in z, we get

$$\left(\rho + \delta \right) \left(\frac{\bar{\alpha}}{\pi v} + 1 - F^\theta(\alpha_l) \right) \Delta \left[\alpha_l \right] = \left(\frac{\bar{\alpha}}{\pi v} + 1 - F^\theta(\alpha_h) \right) \Delta \left[z \right]$$

So an increase in z increases α_h. The effect on the size of the financial sector is ambiguous:

$$\frac{\Delta \left[n \right]}{n} \left(\int_{\alpha_l}^{\alpha_h} m\left(\alpha \right) dF^\theta \left(\alpha \right) + \frac{\mu}{\pi} \right) = m(\alpha_l) f^\theta(\alpha_l) \Delta \left[\alpha_l \right] - \left(F^\theta \left(\alpha_h \right) - F^\theta \left(\alpha_l \right) \right) \Delta \left[z \right]$$

On the one had, it takes more resources to monitor a given set of firms. On the other hand, the pool of firms that are monitored shrinks. Consider now a shift in the function f^θ. We get

$$\left(\rho + \delta \right) \left(\frac{\bar{\alpha}}{\pi v} + 1 - F^\theta \left(\alpha_l \right) \right) \Delta \left[\alpha_l \right] = -\Delta \left[F^\theta \right] (\alpha_l) m(\alpha_l) - \int_{\alpha_l}^{\alpha_h} m\left(\alpha \right) \Delta \left[f^\theta \right] \left(\alpha \right) d\varepsilon$$

and

$$\left(\int_{\alpha_l}^{\alpha_h} m\left(\alpha \right) dF^\theta \left(\alpha \right) + \frac{\mu}{\pi} \right) \frac{\Delta \left[n \right]}{n} = -\int_{\alpha_l}^{\alpha_h} m\left(\alpha \right) \Delta \left[f^\theta \right] \left(\alpha \right) d\varepsilon + m(\alpha_l) f^\theta(\alpha_l) \Delta \left[\alpha_l \right]$$

B.3 Proof of proposition 7

Consider a shift in π

$$\left(\rho + \delta \right) \left(\frac{\bar{\alpha}}{\pi v} + 1 - F^\theta \left(\alpha_l \right) \right) \Delta \left[\alpha_l \right] = \Delta \left[\frac{1}{\pi} \right] \left(m(\alpha_l) \frac{\bar{\alpha}}{v} - \mu \right)$$

so if π goes up, α_l goes up since $m(\alpha_l) \bar{\alpha}/v < \mu$. The effect on n is ambiguous:

$$\left(\int_{\alpha_l}^{\alpha_h} m\left(\alpha \right) dF^\theta \left(\alpha \right) + \frac{\mu}{\pi} \right) \frac{\Delta \left[n \right]}{n} = \frac{1 - n}{n} \Delta \left[\frac{\mu}{\pi} \right] + m(\alpha_l) f^\theta(\alpha_l) \Delta \left[\alpha_l \right]$$

Consider a increase in θ. This increases v, and:

$$\left(\rho + \delta \right) \left(\frac{\bar{\alpha}}{\pi v} + 1 - F^\theta \left(\alpha_l \right) \right) \Delta \left[\alpha_l \right] = m(\alpha_l) \Delta \left[\frac{\bar{\alpha}}{\pi v} \right] - \left(\frac{\bar{\alpha}}{\pi v} + 1 - F^\theta(\alpha_h) \right) \Delta \left[\theta \right]$$

so clearly α_l goes down. The effect on n is ambiguous:

$$\left(\int_{\alpha_l}^{\alpha_h} m\left(\alpha \right) dF^\theta \left(\alpha \right) + \frac{\mu}{\pi} \right) \frac{\Delta \left[n \right]}{n} = \left(F^\theta \left(\alpha_h \right) - F^\theta \left(\alpha_l \right) \right) \Delta \left[\theta \right] + m(\alpha_l) f^\theta(\alpha_l) \Delta \left[\alpha_l \right]$$

Table 1: Data

Period	Finance Share of GDP	Investment Share of Firms with Income<0.33*Capex	Credit Market Instruments over GDP
1955-1965	0.035	0.166	0.373
1966-1975	0.040	0.216	0.473
1976-1985	0.049	0.275	0.524
1986-1995	0.061	0.350	0.604
1996-2005	0.075	0.398	0.624

Notes: Finance Share of GDP is value added of the Finance and Insurance industry divided by GDP, both measured in current dollars. Investment share of low cash firms is the fraction of all capital expenditure in Compustat accounted for by firms whose income is less than a third of their capital expenditures. See also Figure 5. Credit Market Instruments are for the non financial corporate sector. Source: National Income and Product Accounts, Compustat, and Flow of Funds.

Table 2: Counter- Factuals

	Finance Value Added over GDP	Fraction of Constrained Firms
Starting Value (from model)	0.0350	0.1287
Final Value (from model)	0.0685	0.1287
Demand Shift: predicted value without productivity gains in financial sector	0.0641	0.1923
Productivity Gains: predicted value without change in income-growth option correlation	0.0377	0.0924

Figure 1: The Size of the U.S. Financial Sector
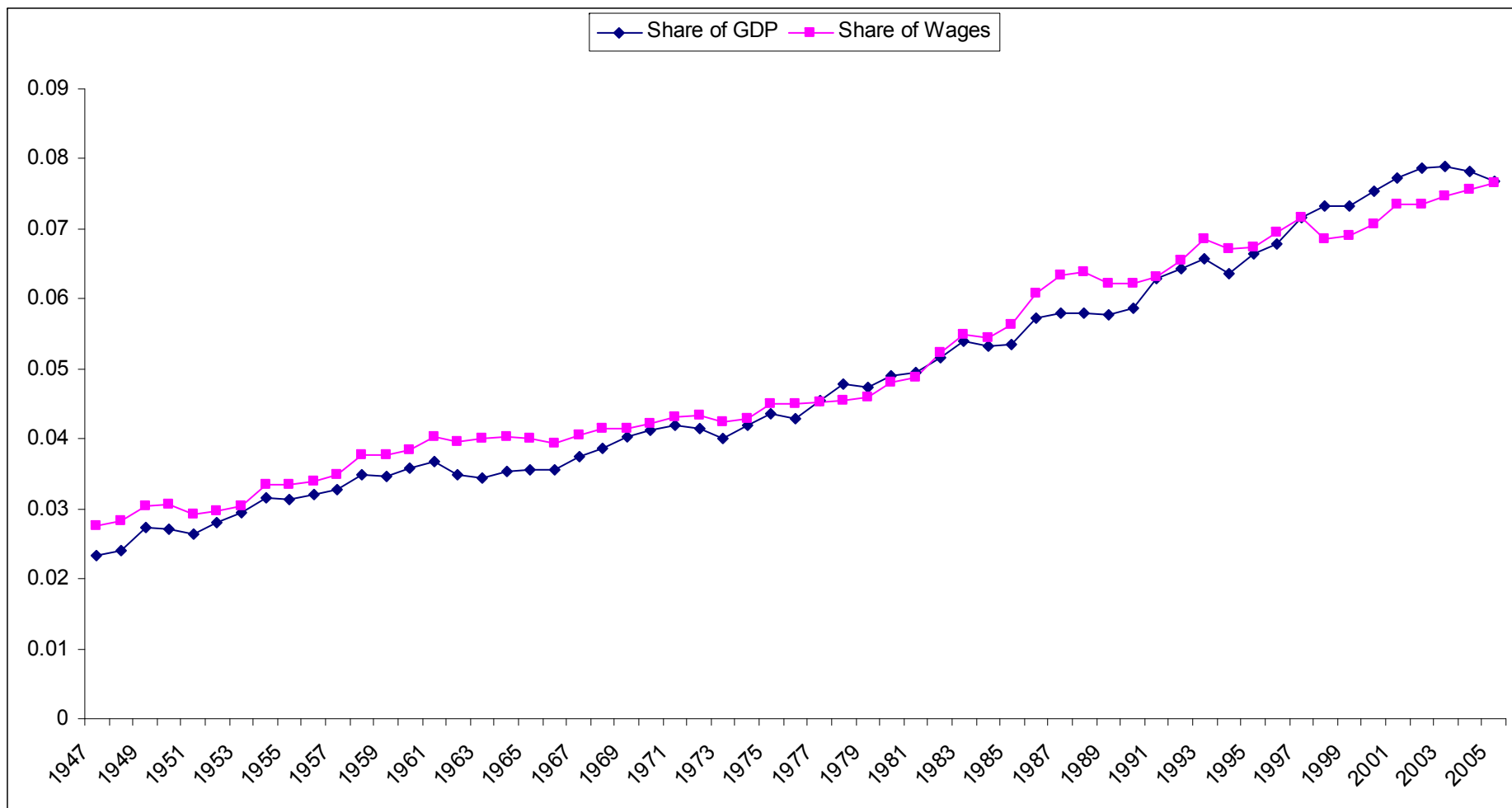
Source: U.S. National Income and Product Accounts

Figure 2: No Moral Hazard
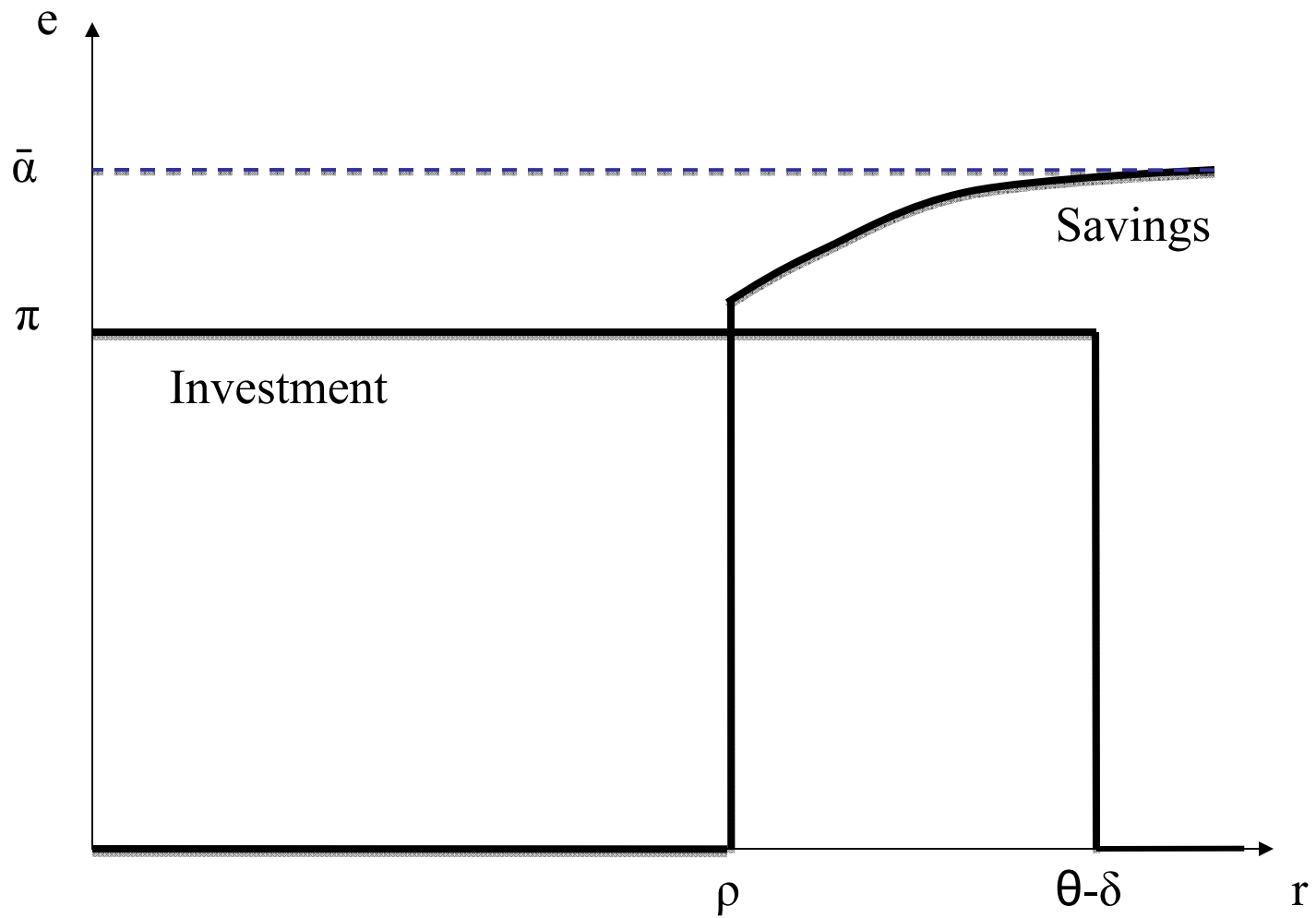
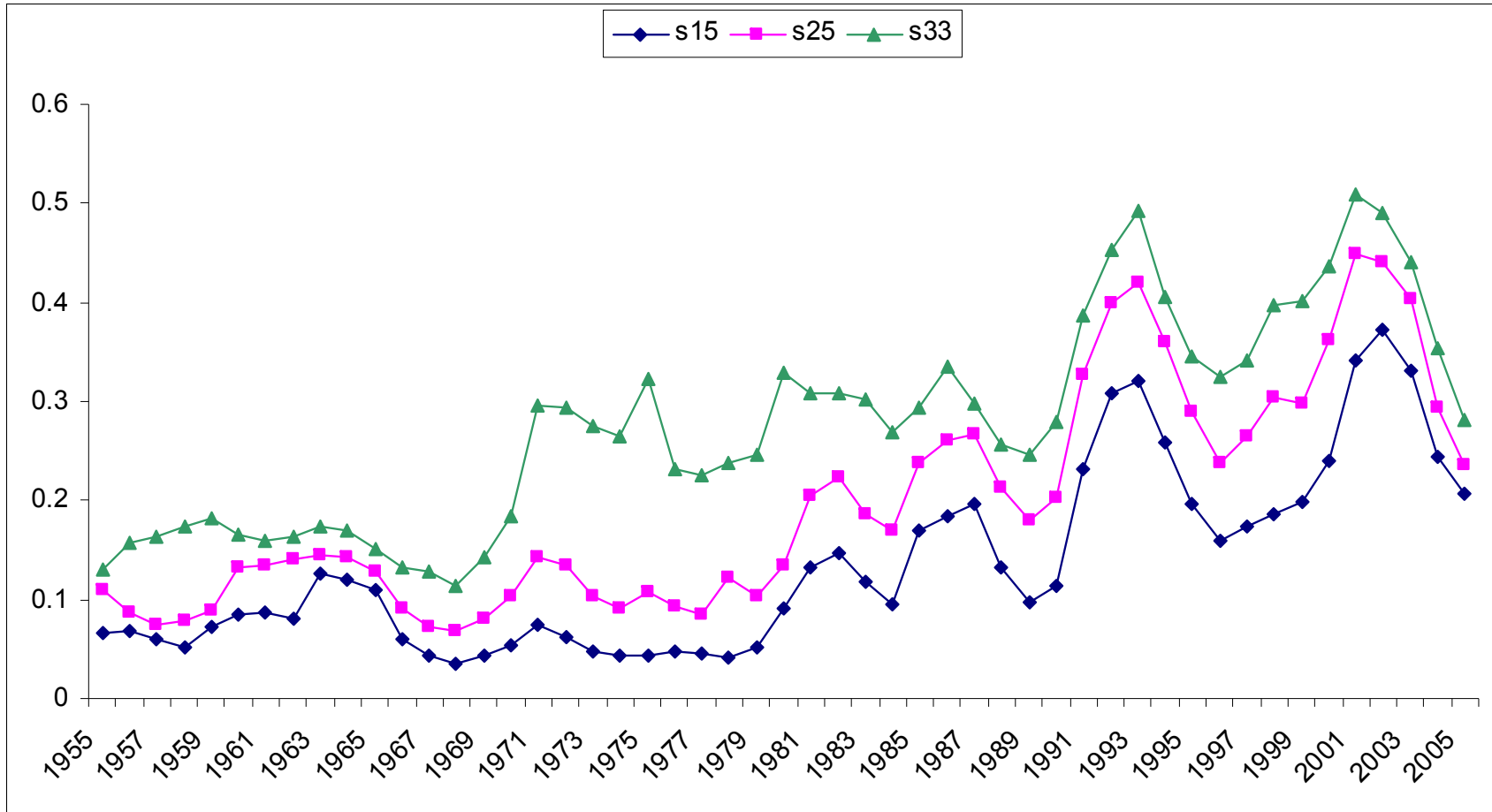Figure 3: Moral Hazard Equilibrium
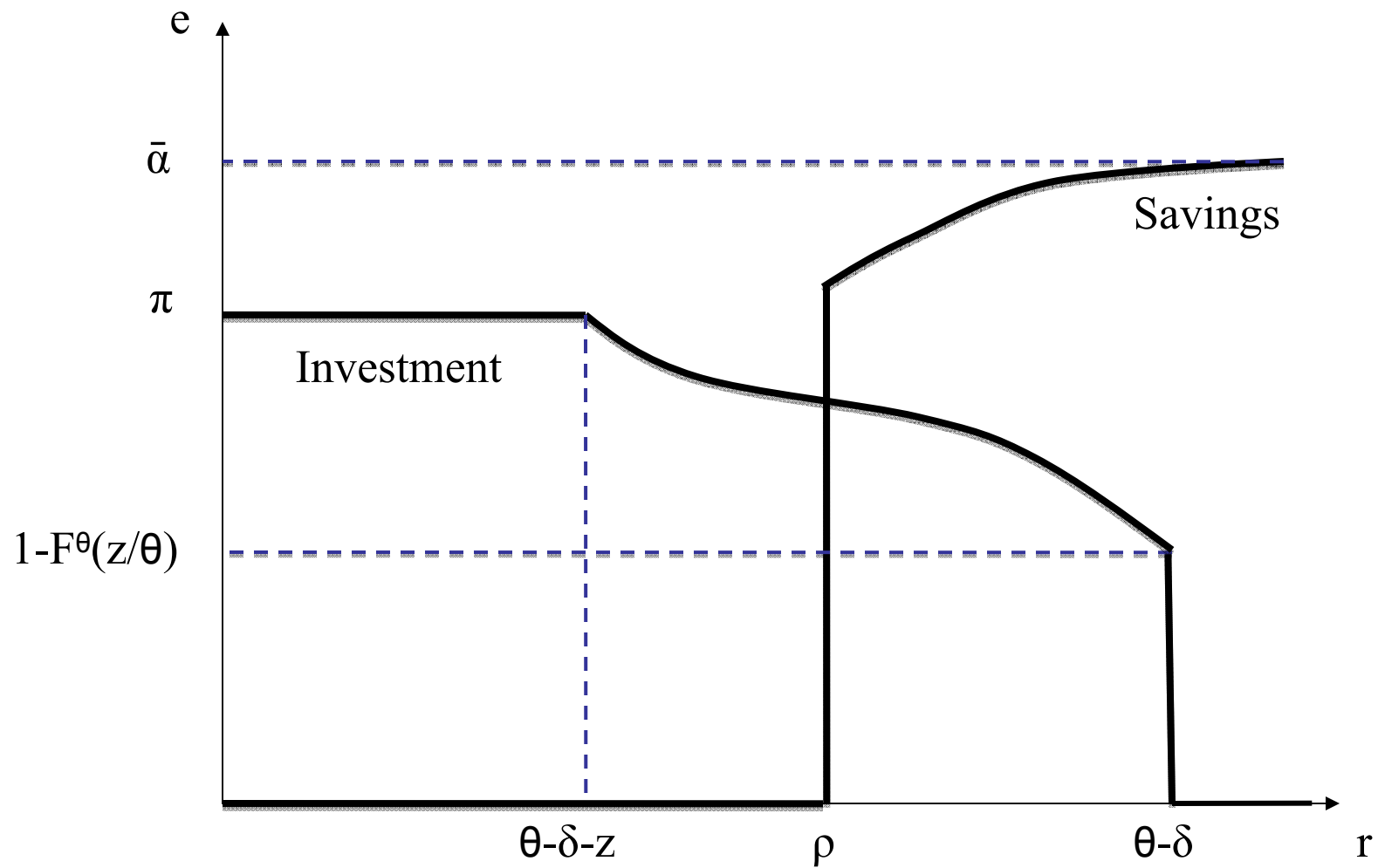
Figure 4: Monitoring Equilibrium
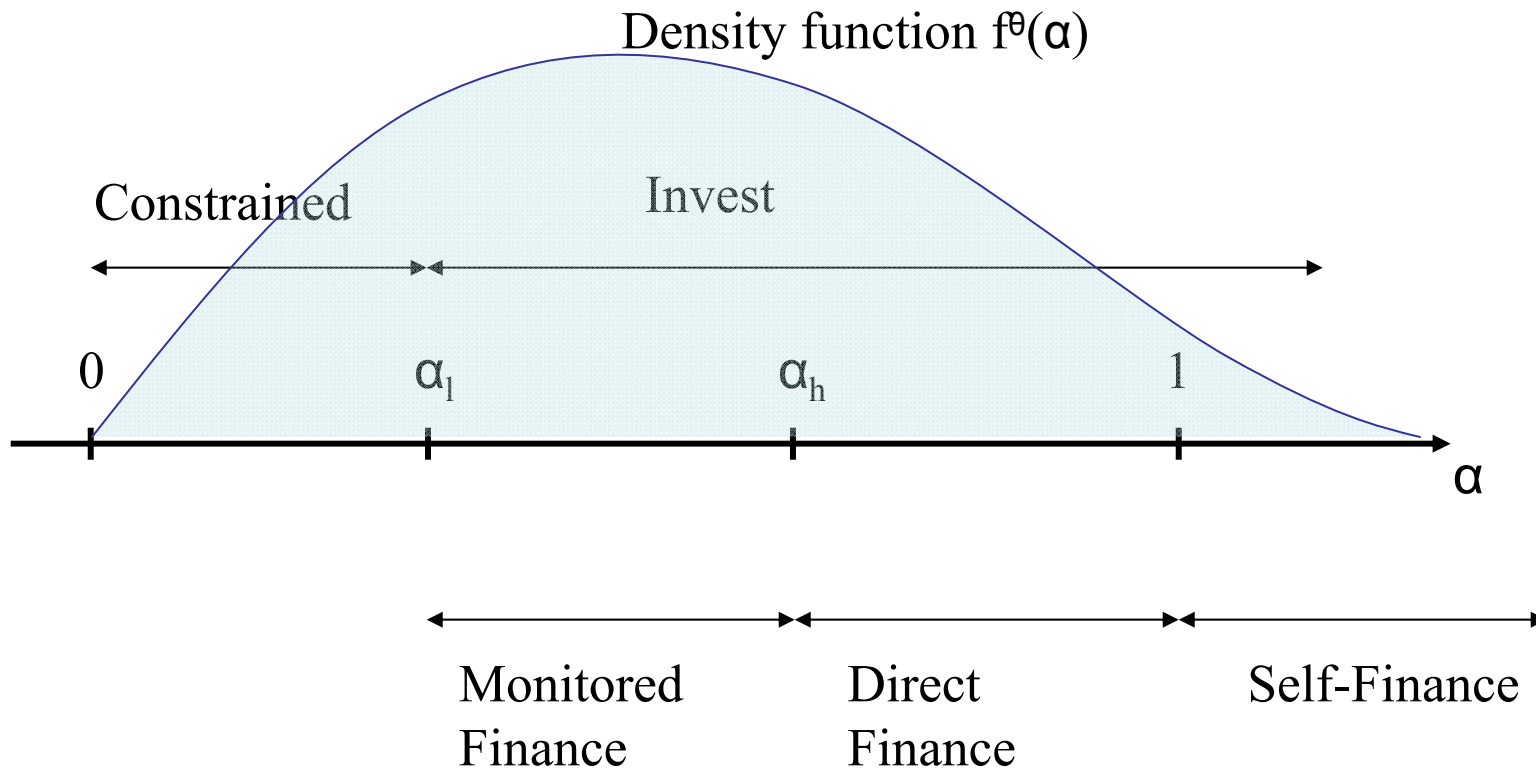
Figure 5: Investment Shares of Low Cash Firms

Notes: Sum of capital expenditures by firms whose income is less than 15%, 25% or 33% of their capital expenditures, divided by the sum of capital expenditures by all the firms in the sample. Sample: Compustat, industrial firms, excluding finance and real estate.
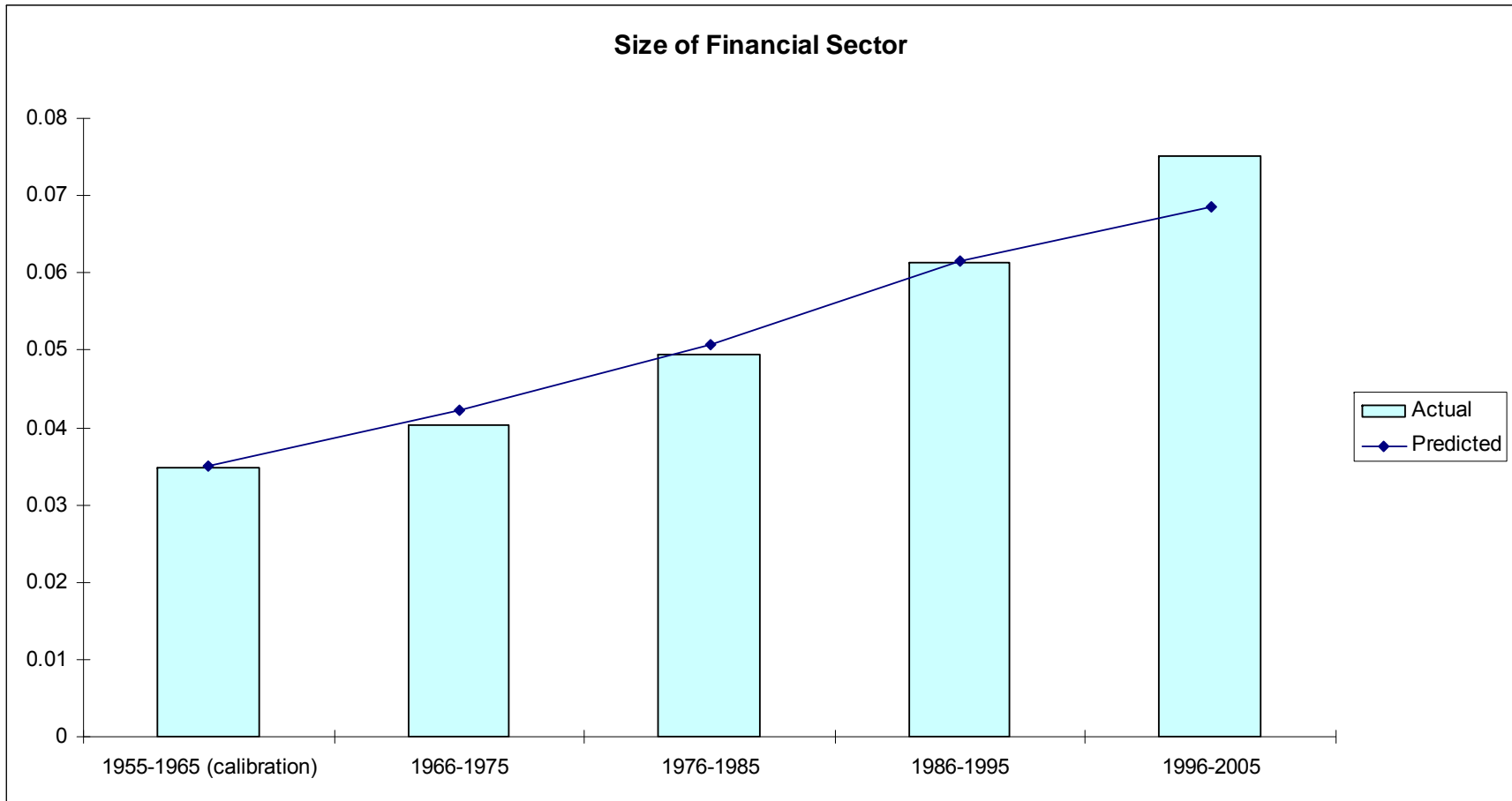
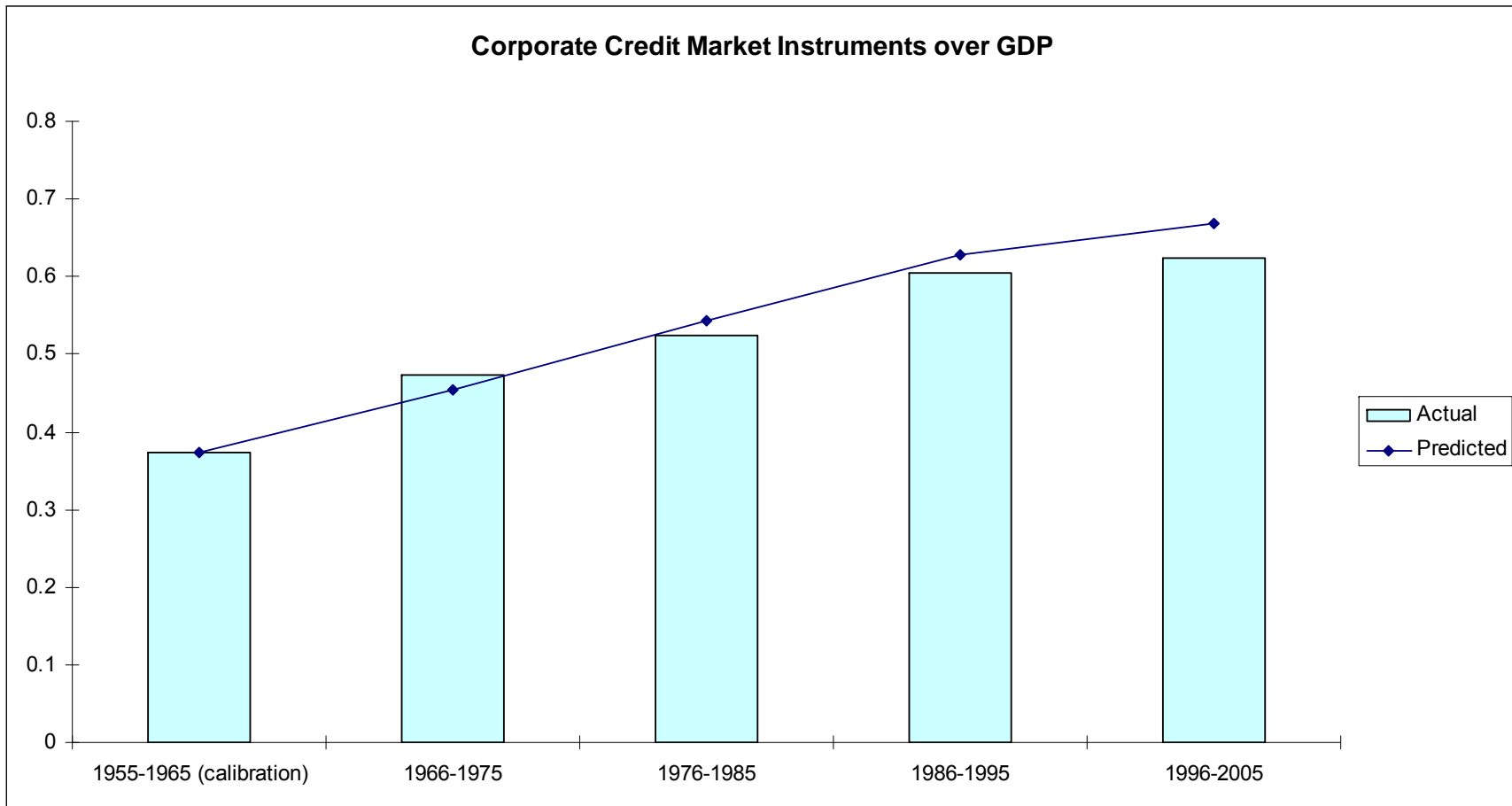# Figure 6: Efficiency Gains in the Financial Sector

SIZE OF FINANCIAL SECTOR

FRACTION OF CONSTRAINED FIRMS
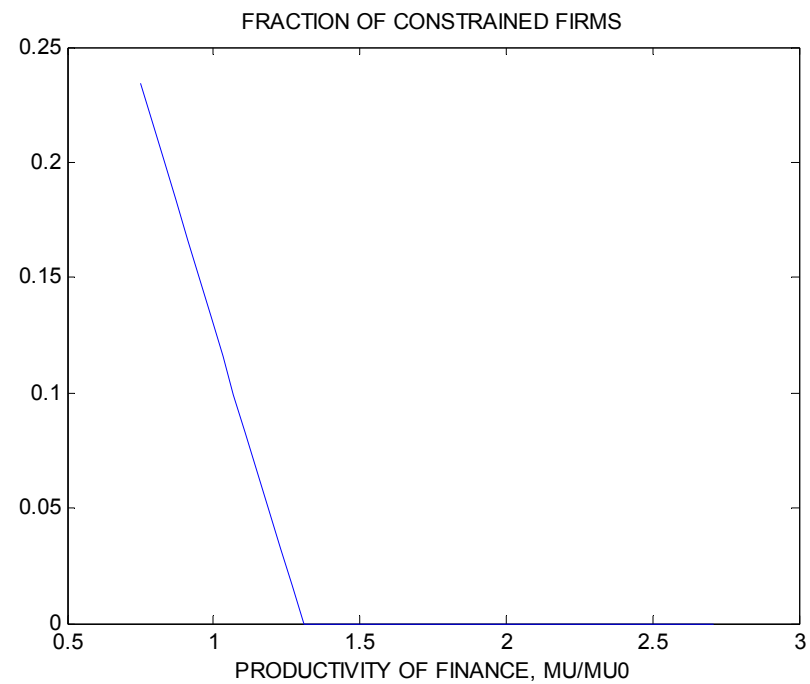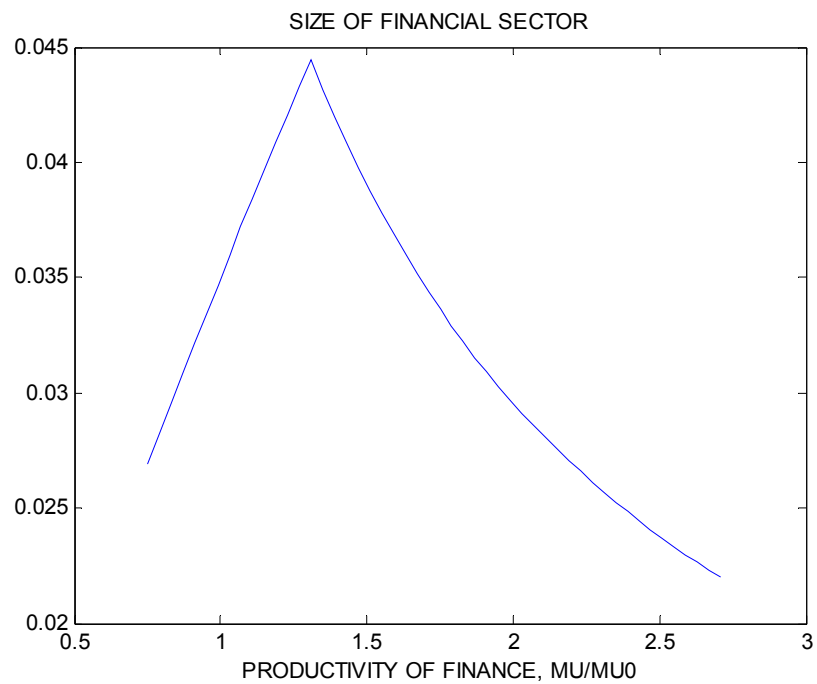
Notes: Model calibrated to the US economy in 1955-65.

Figure 7: Simulation of Calibrated Model

Size of Financial Sector

Notes: The model is calibrated to match the US economy in 1955-65. The simulation assumes that the investment share of GDP is constant. The only input in the simulation is the investment share of low cash firm (s33) from Figure 5. No financial variable is used in the simulation beyond the first period.

Figure 8: Simulation of Calibrated Model

Corporate Credit Market Instruments over GDP

Legend: Actual, Predicted

X-axis categories: 1955-1965 (calibration), 1966-1975, 1976-1985, 1986-1995, 1996-2005

Y-axis: 0 to 0.8

Notes: The model is calibrated to match the US economy in 1955-65. The simulation assumes that the investment share of GDP is constant. The only input in the simulation is the investment share of low cash firm (s33) from Figure 5. No financial variable is used in the simulation beyond the first period.