Has the U.S. Finance Industry Become Less Efficient? On the Theory and Measurement of Financial Intermediation

Thomas Philippon*

February 2012

Abstract

I provide a quantitative interpretation of financial intermediation in the U.S. over 140 years. I measure the cost of intermediation on the one hand, and the production of financial services on the other. I find the following results: (i) intermediation is produced under constant returns to scale; (ii) "quality" adjustment for changes in borrowers' characteristics are important; (iii) the unit cost of intermediation in the U.S. economy has historically been around 2% (i.e., creating and maintaining one dollar of intermediation costs about 2 cents); (iv) surprisingly, however, the unit cost of intermediation is higher today than it was a century ago, and it has increased over the past 30 years. One interpretation is that improvements in information technology may have been cancelled out by increases in other financial activities whose social value is difficult to assess.

STILL PRELIMINARY

*Stern School of Business, New York University; NBER and CEPR. This has been a very long project. The first draft dates back to 2007, with a focus on corporate finance, and without the long term historical evidence. This paper really owes a lot to other people, academics and non-academics alike. I have been fortunate to receive encouraging feedback at various stages of this project from Darrell Duffie, Robert Lucas, Raghuram Rajan, Jose Scheinkman, Robert Shiller, and Andrei Shleifer. I have greatly benefited from the insights of Lewis Alexander, Patrick Bolton, Markus Brunnermeier, John Cochrane, Douglas Diamond, John Geanakoplos, Gary Gorton, Ashley Lester, Andrew Lo, Andrew Metrick, Matthew Rhodes-Kropf, David Robinson, Kenneth Rogoff, Richard Sylla, and Gillian Tett, as well as seminar participants at Stanford, Yale, NYU and the Paris School of Economics. I also thank Paul Krugman for his discussion at the 2011 NY Area Monetary conference. I thank Axelle Ferrière, Peter Gross and Shaojun Zhang for research assistance, and the Smith Richardson Foundation for its financial support.

This paper is concerned with the theory and measurement of financial intermediation. Its contribution is to construct long time series on prices and quantities of intermediation, and to provide a quantitative interpretation of these series. Since the focus is on financial intermediation, the prices are spreads and fees earned by intermediaries, while the quantities are stocks and flows of financial assets and liabilities.

The role of the finance industry is to produce, trade and settle financial contracts that can be used to pool funds, share risks, transfer resources, produce information and provide incentives. Financial intermediaries are compensated for providing these services. The income received by these intermediaries measures the cost of financial intermediation. This income is the sum of all spreads and fees paid by non-financial agents to financial intermediaries, and it is also the sum of all profits and wages in the finance industry. The first contribution of the paper is empirical. I show that the income of financial intermediaries as a share of GDP varies a lot over time. The income share grows from 2% to 6% from 1870 to 1930. It shrinks to less than 4% in 1950, grows slowly to 5% in 1980, and then increases rapidly to more than 8% in 2010. This finding is robust to alternative measures, e.g., excluding net exports of financial services, or scaling by services instead of GDP.

After observing these large historical changes in the finance income share, it is natural to ask the following questions: Is finance a normal good? Should we expect finance to grow with income per capita? How do productivity growth in the non financial sector or technological progress in intermediation affect the size of the finance industry? To answer these questions, I introduce financial services for firms and households in the neoclassical growth model. This is the second contribution of the paper.

Under the assumption of homogenous monitoring (a natural assumptions for monitoring and screening technologies), the model predicts no income effect (i.e., no mechanical tendency for the finance income share to grow with per-capita GDP). The intuition for this result is simple. As borrowers become more productive, the value of monitoring increases even though the monitoring technology itself does not change. Since the opportunity cost of being a banker is the wage in the non-financial sector, and since this wage is proportional to aggregate productivity, the income share of finance remains constant on the balanced growth path. I test this hypothesis and find that it holds well.

As far as the intermediation technology is concerned, we know that efficiency gains lead to a decrease in the finance income share if and only if the elasticity of demand for financial services is less than one. The model relates this elasticity to preferences, production technologies, and shows that it depends crucially on heterogeneity among borrowers. With homogenous borrowers, and for the relevant range of macroeconomic parameters, the elasticity is always less than one. With heterogeneous borrowers and when intermediation is relatively inefficient, the elasticity can be greater than one. The important insight for applied work is that one must distinguish borrowers or projects that are easy to monitor from those that are difficult to monitor. Conceptually, this is akin to performing a quality adjustment on the amount of financial assets that are created.

The third contribution of the paper is to construct a consistent, theory-based measure of output for the finance industry. Conceptually, it is useful to distinguish three types of services:[1]

- (i) Provide liquidity (means of payments, cash management);
- (ii) Transfer funds from savers to borrowers (pools funds, screen and monitor borrowers);
- (iii) Provide information (price signals, advising fees) and/or insurance (diversification, risk management).

Services of type (i) and (ii) involve the creation of various financial assets and liabilities. In the credit market, I measure separately the quantities borrowed by households, farms, non-financial firms, financial firms, and the government. I find that the non-financial corporate credit market is smaller today than it was at its peak of the late 1920s. The most important trends in recent years are the increase in household debt, which exceeds 100% of GDP for the first time in history, and in financial firms' debt, which exceeds that of non-financial firms also for the first time. In the equity market, I measure initial and seasoned offerings. For liquidity I measure deposits, repurchase agreements, and money markets mutual funds. For advising fees I construct a measure of M&A activity. I do not attempt in this paper to measure the informativeness of prices (I discuss the relevant references in Section 6). I then create an output series by aggregating all types of credit, equity issuances, advising and liquidity services produced by the finance industry for the non-financial sector.

The fourth contribution of the paper is to use the theory to perform quality adjustments to the output series. I find that these quality adjustments are important, especially in years where young and risky firms or poor households gain access to financial markets.

Finally I construct the unit cost of intermediation by dividing the total income of the finance industry by my quality adjusted measure of financial intermediation output. This cost has remained roughly stable over the past 130 years. I estimate that it takes between one and a half and two cents to create one dollar of intermediation. However, I also find that the unit cost of intermediation has increased since 1980 and is now significantly higher than it was at the turn of the twentieth century. In other words, the finance industry that sustained the expansion of railroads, steel and chemical industries, and later the electricity and automobile revolutions seems to have been more efficient than the current finance industry. In finance, unlike in other industries (e.g., retail and wholesale trade), the tremendous improvements in information technologies of the past 30 years have not led to a decrease in the average cost of financial intermediation. This might be called the intermediation cost puzzle.

Related literature Financial intermediation does not have a benchmark quantitative framework in the way asset pricing does. By using a model to interpret long time series of prices and quantities, the paper shares the spirit of Mehra and Prescott (1985). It also articulates a puzzle for future research to solve. But because financial intermediation is a more heterogenous field than asset pricing, the paper builds on several strands of literature in finance and monetary economics,

[1] My classification is motivated by the mapping between theory and measurement discussed throughout the paper. It differs a little bit from that Merton (1995).

The first strand is the theory of banking and intermediation. While stylized and focused on macroeconomic predictions, the model developed below is consistent with leading theories of financial intermediation, such as **?**, **?**, Gorton and Pennacchi (1990), Holmström and Tirole (1997), Diamond and Rajan (2001), and Kashyap, Rajan, and Stein (2002). Gorton and Winton (2003) provide a review of the literature on financial intermediation. The focus of this paper differs from this literature in several ways: (i) the measurement of the costs of intermediation; (ii) the simultaneous modeling of household and corporate finance; and (iii) the use of an equilibrium model to interpret the historical evidence.

There is a large literature on financial development, which I cannot discuss here, except to say that it tends to focus on cross-sectional comparisons of countries at relatively early stages of financial development in order to understand the impact of finance on economic growth (e.g. Rajan and Zingales (1998)) and the determinants of financial development (e.g. La Porta, Lopez-de Silanes, Shleifer, and Vishny (1998), Guiso, Sapienza, and Zingales (2004)). It also typically focuses on corporate finance. This paper is more closely related to a recent branch of the literature that seeks to provide risk-adjusted measures of financial productivity (Haldane, Brennan, and Madouros (2010), Basu, Inklaar, and Wang (2011)) and that considers the possibility of inefficient financial development (Glode, Green, and Lowery (2010), Bolton, Santos, and Scheinkman (2011)). Philippon and Reshef (2007) share the historical perspective of this paper, but that paper is purely empirical and focused on the quality and compensation of labor in the finance industry. The large historical changes in the finance share of GDP were first documented and discussed in Philippon (2008), but that paper only focused on corporate credit, which I now estimate to be only one third of output of the finance industry.[2]

In its account of liquidity services provided by the finance industry, the paper is also related to the classic literature on money and banking. Lucas (2000) provides a benchmark analysis of money demand. A recent branch of this literature has focused on the rise of market-based intermediation, also called shadow banking. Pozsar, Adrian, Ashcraft, and Boesky (2010) document the structure of shadow banking. Gorton and Metrick (2012), Stein (2012), and Gennaioli, Shleifer, and Vishny (2011) emphasize the importance of investors demand for risk free assets. Gorton, Lewellen, and Metrick (2012) argue that much of the shadow banking activity (pooling and tranching) happens to satisfy this demand for risk free assets. I attempt to account for these activities by measuring shadow deposits, such as money market mutual funds and repurchase agreements. The rise of shadow banking also diminishes the relevance of the traditional literature focused on efficiency in banking. That literature did provide measures of productivity in banking (see Wang, Basu, and Fernald (2008) for a discussion), but it focused on net interest income, which is only about half of the income of today's large banks (see the numbers for JP Morgan in the appendix). An important point developed below is that is it difficult to break down the income earned by the finance industry into economically meaningful components.

[2]The paper did not consider household credit, and did not account for liquidity services, which have become important with the rise of the shadow banking system. Another, more technical issue is that the production functions in Philippon (2008) were not appropriate, as is now clear from Proposition 1.

Finally, it is important to emphasize an important limitation of my analysis: it does not deal with financial crises and risk taking. For instance, my output series include all corporate borrowing by Telecom companies in the late 1990s and all subprime and home equity borrowing by households in the mid 2000s. In doing so, I never ask whether borrowing is appropriate or excessive, and I therefore miss the crucial insights of Reinhart and Rogoff (2009). Similarly, I consolidate the earnings of financial intermediaries without controlling for systemic risk taking.[3]

The remaining of the paper is organized as follows. In Section 1, I construct my measure of the cost of financial intermediation. Sections 2 and 3 present benchmark models of corporate and household finance to organize the discussion. Section 4 presents measures of output for the finance industry. In section 6, I discuss the role of information technology, price informativeness, financial derivatives, risk sharing, and trading. Section 7 concludes and discusses avenues for future research.

1 Income Share of Finance Industry

In this section, I present the first main empirical fact: the evolution of the total cost of financial intermediation in the US over the past 140 years. As argued in the introduction, there is no clear way to break down the income earned by the finance industry into economically meaningful components. Insurance companies and pension funds perform credit analysis, fixed income trading provides liquidity to credit markets, and securitization severs the links between assets held and assets originated. These issues are compounded by regulatory changes in the range of activities that certain intermediaries can provide. I therefore focus on a consolidated measure of income, the sum of all interest and non interest income earned by all financial intermediaries.

1.1 Raw Data[4]

There are various ways to define the size of the financial sector. Conceptually, the measure is

$$\phi = \frac{Finance\,Income}{Total\,Income}$$

There are two immediate issues:

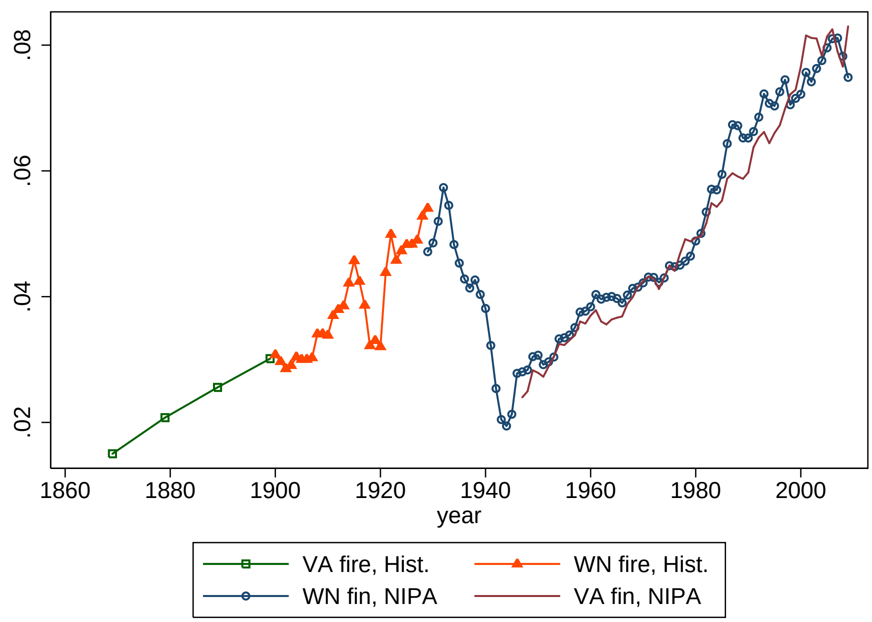
- Definition of "Finance." For the most part, financial activities are classified consistently over time (but subsectors within finance are not). The main issue is with real estate. The value added of the "real estate" industry includes rents and imputed rents for home owners. Whenever possible, I exclude real estate. In my notations, all variables indexed with "fin" include finance and insurance and exclude real estate. This is not possible before 1929. In this case I use the compensation of employees whenever possible.

[3] See for instance Adrian and Shin (2008), Krishnamurthy (2009), and Acharya, Pedersen, Philippon, and Richardson (2009) for recent discussions.

[4] The paper uses a lot of data sources. To save space in the paper, all the details regarding the construction of the series are provided in a separate online appendix.

- Definition of "Income." The best measure in theory is total income. In this case, ϕ is the nominal income of the finance industry over the GDP of the US economy. However, this is only acceptable if we can measure the finance industry without imputed rents from the real estate sector. When this is not possible, a good alternative is to use the compensation of employees because the share of real estate is small. In this case, ϕ is the compensation of employees in finance over the total compensation of employees in the US. For the post-war period, the two measures display the same trends, even though annual changes can differ. This simply means that, in the long run, the labor share in the finance industry is the same as the labor share in the rest of the economy. In the short run, of course, profit rates can vary.

Figure 1: Income Share of Finance (non-farm civilian)



Notes: VA is value added, WN is compensation of employees, "fin" means finance and insurance, "fire" means finance, insurance, and real estate. For "NIPA", the data source is the BEA, and for "Hist" the source is the Historical Statistics of the United States.

Figure 1 displays various measures of the share of the Finance and Insurance industry in the GDP of the United States estimated from 1870 to 2009. For the period 1947-2009, I use value added and compensation measures from the Annual Industry Accounts of the United States, published by the Bureau of Economic Analysis (BEA). For 1929-1947, I use the share of employee compensation because value added measures are either unavailable or unreliable. For 1870-1929 I use the Historical Statistics of the United States.[5] More detail regarding the various data sources can be found in Philippon and Reshef (2007) and in the Data Appendix (available online).

The first important point to notice is that the measures are qualitatively and quantitatively consistent. It is thus possible to create one "extended" series simply by appending the older data to the newer ones.[6] The second

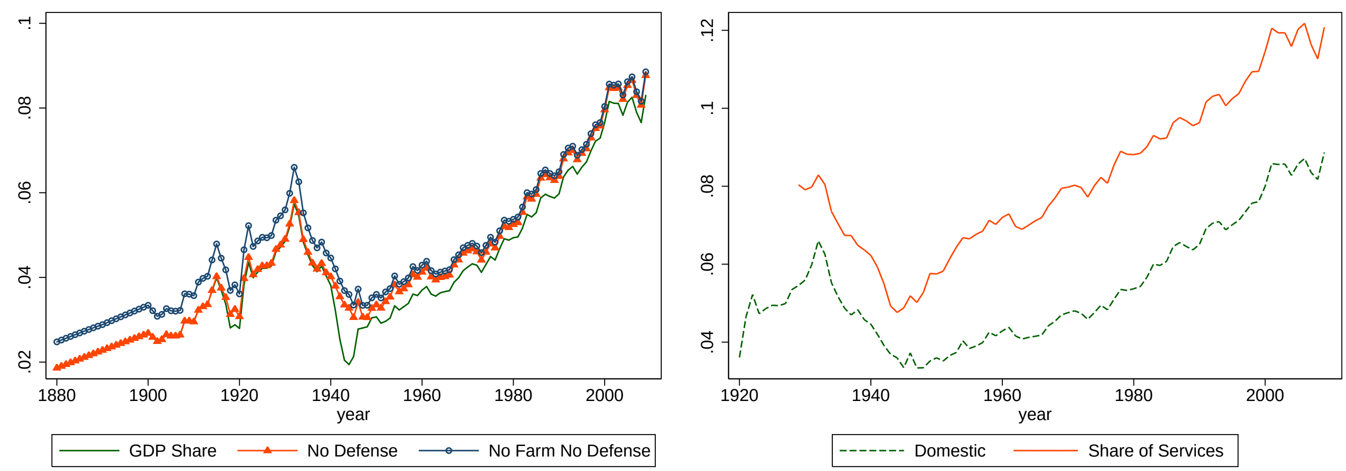
[5] Carter, Gartner, Haines, Olmstead, Sutch, and Wright (2006).
[6] Other measures based on Martin (1939) and Kuznets (1941) give also give consistent values.

key point is that finance was smaller in 1980 than in 1925. Given the outstanding real growth over this period, it means that finance size is not simply driven by income per capita.

1.2 Adjusted Measures

Before discussing theoretical interpretations it is useful to present adjusted series and consider the impact of globalization and the rise in services.

Figure 2: Income Share of Finance (alternative measures)



Notes: GDP Share is the Income of the Finance Industry divided by GDP, constructed from the series in Figure 1. "No Defense" uses GDP minus defense spending, and "No Farm No Defense" uses non-farm GDP minus defense spending. Domestic shares excludes net exports of finance and insurance companies. Share of Services uses the BEA definition of services.

War and Structural Change

During peace time and without structural change, it would make sense to simply use GDP as the relevant measure of total income. Two factors can complicate the analysis, however. First, WWI and WWII take resources away from the normal production of goods and services. Financial intermediation should be compared to the non-war related GDP. To do so, I construct a measure of GDP excluding defense spending. The second issue is the decline in farming. Since modern finance is related to trade and industrial development, it is also useful to estimate the share of finance in non-farm GDP.

The left panel of Figure 2 presents the finance share of non-defense GDP, and of non-farm, non-defense GDP (or compensation, as explained above). Both adjustments make the series more stationary. In particular, using non-defense GDP removes the spurious temporary drop in the unadjusted series during WWII.

I use the defense adjusted share as my main measure. The share of finance starts just below 2% in 1880. It reaches a first peak of almost 6% of GDP in 1932. Note that this peak occurs during the Great Depression, not in 1929. Between 1929 and 1932 nominal GDP shrinks, but the need to deal with rising default rates and to restructure corporate and household balance sheets keeps financiers busy. Similarly, the post-war peak occurs not in 2007 but in 2010, just below 9% of GDP.

Finance versus Services

Is finance different from other service industries? Yes. The right panel of Figure 2 also plots the share of finance in service GDP. It is of course (mechanically) higher than it is in total GDP, but the pattern is the same (the other fast growing service industry is health care, but it does not share the U-shaped evolution of Finance from 1927 to 2009).

Globalization and Trade in Financial Services

In Figure 1, I divide by US GDP. This makes sense if financial services are produced and used locally. But in the recent part of the sample, the US presumably exports some investment banking services abroad. It turns out, however, that this adjustment is small.

The right panel of Figure 2 displays the ratio of income minus net exports for finance over non-farm civilian GDP. The figure is almost identical to the previous one. The reason is that the U.S., unlike the U.K. for instance, is not a large exporter of financial services. According to IMF statistics, in 2004, the U.K. financial services trade balance was +\$37.4 billions while the U.S. balance was -\$2.3 billions: the U.S. was actually a net importer. In 2005, the U.K. balance was +\$34.9 billions, and the U.S. balance was +\$1.1 billions.[7]

The timing is also different. Estevadeordal, Frantz, and Taylor (2003) show that the period 1870-1913 marked the birth of the first era of trade globalization and the period 1914-1939 its death. The period 1918 to 1930 is the first large scale increase in the size of the finance industry, precisely as globalization, measured by the ratio of trade to output, was receding. For the more recent data, Obstfeld and Taylor (2002), and Bekaert, Harvey, and Lumsdaine (2002) show that financial globalization happens relatively late, in the 1990s, while Figure 1 shows that the growth of the financial sector accelerates around 1980. Globalization therefore does not account for the evolution of the U.S. financial sector.

The goal of the next two sections is to build simple models that can shed light on the following questions: Is finance a normal good? Should we expect finance to grow with income? How does productivity in the non finance sector affect the size of finance? What should be the impact of technological progress in finance on the size of finance?

To answer these questions, I introduce financial services to firms and households in the neoclassical growth model.[8] Section 2 focuses on corporate finance, while Section 3 focuses on household finance.

[7]There is, of course, some trade within the financial sector, notably between the U.S. and the U.K., but the growth in the GDP share of finance is not due to large net exports.

[8]The Neoclassical growth model can easily be extended to accommodate two sectors. It is well known that the properties of this model depend on the elasticity of substitution between the two sectors (Baumol (1967)). The nominal GDP share of sector i increases with relative technological progress in sector i if and only if the elasticity of substitution is less than one. I argue, however, that the traditional multi-sector model is not useful to analyze financial intermediation because it is not the reduced form of any sensible model

2 Corporate Finance Theory

There is a long tradition of modeling corporate financial services. I do not attempt to do justice to this rich corporate finance literature. Rather, I highlight the macroeconomic implication of technological progress in the finance industry on the size of credit markets and the GDP share of the industry.

I use the neoclassical growth model as a benchmark. Since it is well known, the details are in the Appendix. Output is produced with constant returns technology $Y_t = F(A_t \bar{n}_t, K_t)$. There is a representative household who makes all the inter-temporal decisions. She owns the capital stock K_t, which depreciates at rate δ, and she maximizes her expected lifetime utility $\mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t u(C_t)\right]$. The household has preferences $u(C) = \frac{C^{1-\rho}}{1-\rho}$ with constant relative risk aversion ρ, and inelastic labor supply normalized to $\bar{n} = 1$. The economy is non-stationary. The driving force is the labor-augmenting technology shock $A_t = (1 + \gamma_t) A_{t-1}$. I use the convention of upper-case letters for variables with trends, and lower-case letters for their de-trended counterparts. For instance, for capital I write $k_t \equiv \frac{K_t}{A_t}$ and for consumption $c_t \equiv \frac{C_t}{A_t}$. Given constant returns, we can write output as $Y_t = A_t F(\bar{n}_t, k_t)$. I focus on the balanced growth path with constant growth rate γ. Let r be the interest rate received by savers. The Euler equation of consumer pins down the equilibrium rate on the balanced growth path:

$$r = \beta^{-1} (1 + \gamma)^{\rho} - 1. \tag{1}$$

The key questions addressed in this section are:

- What determines the demand for intermediation services (i.e., what is the value of intermediation for the non-financial sector)?
- How are these services produced (i.e., what are the characteristics of the production function)?

I proceed as follows. I start from the simplest specification where borrowers are homogenous, all capital is intermediated, and financial services are produced linearly from final goods. This provides a tight characterization of equilibrium intermediation. I then introduce heterogeneity, which is critical to understand quality adjustments. Finally, I generalize the production function.

2.1 Homogenous Borrowers

Let ψ be the cost of financial intermediation per unit of asset.[9] With competitive intermediation, ψ is also the price of intermediation. I start by assuming that all capital is intermediated. Industrial firms therefore solve the following program $\max_{n,k} F(n,k) - (r + \delta + \psi) k - wn$, where w is the productivity adjusted wage: $w \equiv W_t / A_t$.

of financial intermediation (more details can be found in the appendix). This will become evident in the next two sections. Section 2 introduces the simplest model of intermediation services to firms. In this model, the elasticity depends both on the shape of the distribution of borrowers and on the efficiency of the supply of financial services. Section 3 introduces financial services to households, and shows again that the standard multi-sector model is not useful to understand the finance industry. Instead, we must explicitly model financial intermediation.

[9] The Appendix describes a simple monitoring model to justify the use of financial intermediaries.

The model is summarized by three equations. The first is the Euler equation (1). The second is the capital demand equates the marginal product of capital to its user cost:

$$\frac{\partial F}{\partial k}(1,k) = r + \delta + \psi. \tag{2}$$

For future reference, we define the neoclassical level of capital k^* by $\frac{\partial F}{\partial k}(1,k^*) \equiv r + \delta$. Detrended output is simply $y = F(1,k)$ since $\bar{n} = 1$ in equilibrium, and we have the capital accumulation/resource constraint

$$(\gamma + \delta + \psi)\, k = F(1,k) - c. \tag{3}$$

On the balanced growth path the system is block diagonal since (1) pins down r, then (2) pins down the capital stock. The consumption $c = w + (r - \gamma)\, k$ is labor income plus capital income in excess of growth.[10] The finance share of GDP is

$$\phi = \frac{\psi k}{F(1,k)} \tag{4}$$

In this model the real output of the finance industry is proportional to k, since each unit of k requires one unit of monitoring. The parameter ψ is the unit cost of monitoring, so ψk is the nominal output of the financial sector. There is an elasticity of demand because the capital output ratio decreases with ψ. The following Proposition summarizes the prediction of the model regarding the impact of improvement in financial intermediation.

Proposition 1. *In the homogenous-borrowers model, improvements in financial intermediation (lower ψ) lead to an increase in capital, real wages, and consumption. The GDP share of the finance industry decreases as long as*

$$(\sigma - 1)\, \psi < r + \delta, \tag{5}$$

where σ is the elasticity of substitution between capital and labor.

Proof. The fact that k increases when ψ decreases is clear from (2). The impact on w is clear, and $c = w + (r - \gamma)\, k$ increases because both w and k increase. Let us use a production function with a constant elasticity of substitution to understand the behavior of the GDP share of finance. Output is given by $y = F(n,k) = \left(\alpha \bar{n}^{\frac{\sigma-1}{\sigma}} + (1-\alpha)\, k^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$, where α is the labor share and σ is the elasticity of substitution between capital and labor. In the Appendix, I show that $\frac{d \log \phi}{d\psi} = \frac{1}{\psi} - \frac{\sigma}{r+\delta+\psi}$, which proves the result. $\square$

An important special case is one of Cobb-Douglass technology. In this case, better finance always leads to smaller finance since $\sigma = 1$. One can easily obtain closed form solutions. Capital is simply $k = \left(\frac{1-\alpha}{r+\delta+\psi}\right)^{\frac{1}{\alpha}}$ and output is $y = k^{1-\alpha}$. The consumption output ratio is $\frac{c}{y} = \alpha + (r - \gamma)\frac{k}{y}$ and we know that $\frac{k}{y} = \frac{r-\gamma}{r+\delta+\psi}$. If ψ goes

[10] This comes from the resource constraint (3):$c = f(k) - (\gamma + \delta + \psi)\, k = f(k) - (r + \delta + \psi)\, k + (r - \gamma)\, k$ and the fact that $wn = f(k) - (r + \delta + \psi)\, k$ and $n = 1$.

down, k/y goes up, so c/y goes up (but consumption over capital goes down). With Cobb-Douglass technology, equation (4) simply becomes

$$\phi = (1-\alpha)\,\frac{\psi}{r+\delta+\psi}$$

and it is easy to see that the GDP share decreases when the intermediation cost ψ decreases.

What are the quantitative implications of the model? Most empirical macroeconomic studies use the Cobb-Douglass technology. In this case the model predicts that improvement in financial intermediation should lead to a lower share of finance in GDP. We can go further and ask how high σ would need to be to overturn this prediction. In the empirical section, I find ψ in the range of 1 to 2%. Let us take 2% to be conservative. Standard estimates of the real interest rate are around 1-3%, and of depreciation around 5-10%. Let us take $r+\delta$ of 10%. To violate the condition, we would need $\sigma > 1 + 10/2 = 6$. This is outside the range of plausible elasticities. Most macroeconomic estimates suggest that σ is close to 1. Chirinko, Fazzari, and Meyer (2011) estimate less than 1 for the U.S., while Raurich, Sala, and Sorolla (2011) estimate ranges of 0.6-0.9 for the U.S. and 1.2-1.6 for Spain. It is therefore safe to conclude that condition (5) is satisfied.[11]

2.2 Heterogenous Monitoring Needs

The homogenous borrower model described above is a useful benchmark, but it fails to capture some important features of corporate finance. To give an extreme example, corporate finance involves issuing commercial paper for blue chip companies as well as raising equity for high-technology start-ups. It is plainly obvious that the monitoring requirements per unit of intermediated funding are vastly different. This will play an important role when we match the model to the data. The homogenous borrower model also fails to capture the idea that financial development gives access to credit to borrowers who would otherwise be shut out from the markets. As we will see, modeling this feature is important when thinking about the GDP share of finance, technological progress, and shocks to credit demand.

Let us therefore consider a model with heterogeneity and decreasing returns at the firm level.[12] Let k be the (endogenous) number of firms, and let $n = \bar{n}/k$ be employment per-firm. Each firm operates A_t units of capital and hires n workers to produce $A_t f(n)$ units of output, where f is increasing and concave. Decreasing returns come

[11] What is a plausible interpretation of the homogenous borrower model? The key assumption is that all financial flows are intermediated at the same cost. This model therefore applies either to an economy where firms are fairly homogenous, or, more realistically, it applies to the part of intermediation services that are required by all borrowers. A good example would be passively managed mutual funds. They provide a cheap way for households to hold diversified portfolios of stocks and bonds. Progress in information technology has lowered the cost per unit of asset held. And they are used by (almost) all households and (almost) all firms. In this case the model predicts that mutual funds should increase the stock of corporate assets (measured at market value) but at the same time decrease the GDP share of intermediation. Funds, trusts and other financial vehicles account for only approximately 0.3% of GDP, which is small relative to the more than 8% share of GDP for finance and insurance as a whole.

[12] Decreasing returns in production are required to make room for heterogeneity since with constant returns borrowers that have even a slight financial disadvantage would not be able to enter.

from the fact that capital is fixed at the firm level. Firms choose employment to maximize (detrended) net income

$$\pi(w) \equiv \max_n f(n) - w_t n$$

Each firm hires n employees such that $f'(n) = w$. The labor demand schedule is the decreasing function $n(w)$.

Macroeconomic adjustment to the stock of capital takes place at the extensive margin, i.e., by firms' entry (and exit) decisions. Firms differ in their need for intermediation services, characterized by the monitoring requirement μ per unit of capital. The mass of potential firms with monitoring requirements below μ is $G(\mu)$. Let us define $G_0 \equiv G(0)$ as the mass of (potential) firms that do not require intermediation services.[13] It is convenient to define the density $g(\mu)$ for $\mu > 0$. We can then write $G(\mu) = G_0 + \int_0^\mu g(x)\,dx$. I assume that there are enough potential entrants to have an interior solution. Recall that we have defined k^* as the neoclassical level of detrended capital. I assume that

$$G_0 \leq k^* < G(\infty).$$

Finally, and as in the homogenous borrower model, I let ψ be the unit cost of monitoring services.

This model is rich enough to capture shifts in demand and supply in the market for financial intermediation. I will use G_0 as an inverse measure of demand shocks and ψ as an inverse measure of supply shocks. To solve the equilibrium, we need to consider firms' entry decisions, and then clear the labor market. Consider entry decisions first. The profits of a firm with monitoring requirements μ are $A_t\pi(w) - (r + \delta + \psi\mu) A_t$. The marginal firm $\hat{\mu}$ is defined by:

$$\pi(w) = r + \delta + \psi\hat{\mu}. \tag{6}$$

Firms above the cutoff $\hat{\mu}$ do not enter. The number of firms, which is also the detrended capital stock, is

$$k = G(\hat{\mu}).$$

Households' preferences are unchanged but their budget constraint becomes $K_{t+1} + C_t \leq (1 + r_t) K_t + W_t + \Pi_t$, where Π_t are aggregate corporate profits, and the aggregate capital is simply $K_t = A_t k$. The Euler equation (1) is unchanged, therefore r is still given by preferences (on the balanced growth path). Since there are k firms, and each employs $n(w)$ workers, clearing the labor market requires $\bar{n} = kn(w) = 1$. From (6), we then get the second

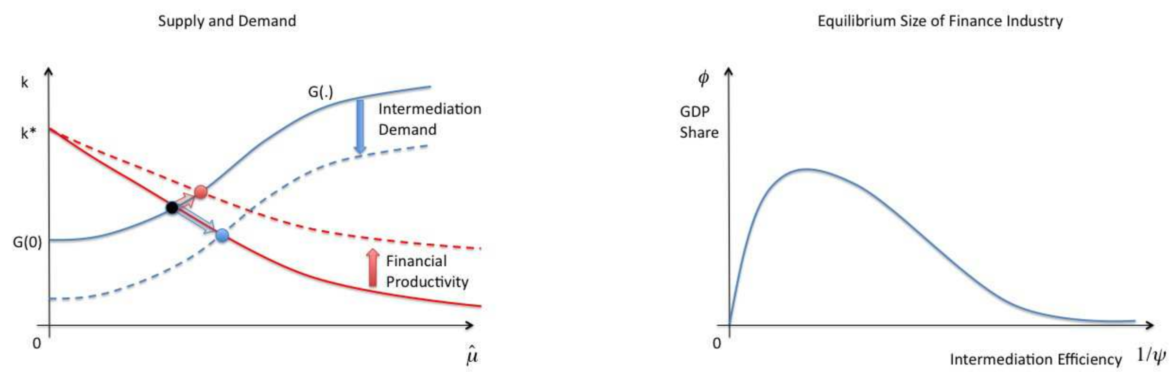
[13] A straightforward extension is to connect the distribution of monitoring needs G to the correlation between cash flows and investment opportunities. When the correlation is high, firms with investment opportunities also have high cash flows and there is no need for intermediation (this corresponds to a high value of G_0). The demand for corporate finance services increases when the correlation decreases. Philippon (2008) provides such a model. From an equilibrium perspective, there are two advantages of this type of micro-foundation. First, the model with moral hazard introduces a new state variable in the system, namely the correlation between cash flows and investment opportunities. Since cash flows are observable, it is then possible (in theory at least) to perform a model-based decomposition of supply and demand shifts. Second, in a model with adverse selection, it is possible to discuss equilibria where the economy is not constrained-efficient and there is room for regulation. These issues are not the main focus of this paper, however.

equilibrium condition

$$G_0 + \int_0^{\hat{\mu}} g(x)\, dx = \frac{1}{n(w)}. \tag{7}$$

The general equilibrium is the solution $(w, \hat{\mu})$ to the system of equations (6) and (7). The equilibrium is depicted on Figure 3.

Figure 3: Equilibrium Corporate Finance



This model nests the homogenous borrower case with a vertical demand curve at $\hat{\mu} = 1$ and the neoclassical growth model with $\hat{\mu} = 0$. The finance share of GDP is now equal to

$$\phi = \frac{\psi}{y} \int_0^{\hat{\mu}} \mu g(\mu)\, d\mu. \tag{8}$$

We have the following proposition:

Proposition 2. *In the heterogenous borrowers model, improvements in financial intermediation (a decrease in ψ) increase the wage and the number of firms but have an ambiguous impact on the GDP share of finance. A higher demand for intermediation (a decrease in G_0) decreases the wage and the number of firms and increases the GDP share of finance.*

Proof. The RHS of (7), $\frac{1}{n(w)}$, is increasing in w. Since $\pi(w)$ is decreasing, we see from (8) that w is decreasing in $\hat{\mu}$. The RHS is therefore a downward schedule in $\hat{\mu}$ that is steeper when the marginal cost ψ is higher. The LHS is obviously an increasing schedule. The equilibrium is unique and depicted in Figure 3. A decrease in ψ increases $\hat{\mu}$, therefore $k = G(\hat{\mu})$ and therefore w from $kn(w) = 1$. The fact that ϕ decreases with G_0 is also clear. Suppose G_0 goes up, then $\hat{\mu}$ goes down while k, and therefore y, go up. From (8), ϕ goes down. $\square$

In general, we see that ϕ is non monotonic in ψ. This property is intuitive. When finance is very inefficient (ψ is very high), all would-be users are priced out: $\hat{\mu}_t$ is 0 and so is ϕ, and only G_0 firms enter. Starting from this level, an improvement in financial intermediation must increase the GDP share of finance. The GDP share of finance reaches its maximum for intermediate levels of financial development. When finance is fully efficient (ψ goes to zero), we get the Walrasian benchmark with equality of lending and borrowing rates and ϕ tends to zero.

With Cobb-Douglass technology, we get net income $\pi(w) = (1-\alpha)\left(\frac{\alpha}{w}\right)^{\frac{\alpha}{1-\alpha}}$, labor demand $n = \left(\frac{\alpha}{w}\right)^{\frac{1}{1-\alpha}}$. Therefore $k_t = \left(\frac{\alpha}{w_t}\right)^{\frac{-1}{1-\alpha}}$ and we can substitute in the equilibrium condition (7) to get the simple condition $G_0 + \int_0^{\hat{\mu}} g(x)\, dx = \left(\frac{1-\alpha}{r+\delta+\psi\hat{\mu}}\right)^{\frac{1}{\alpha}}$.

2.3 Production of Financial Services

A simplification of the above monitoring technology is that monitoring services are produced directly from final goods with a constant marginal cost ψ.[14] An obvious extension is to consider the more realistic case where financial services are produced with capital and labor instead. This is rather straightforward and leaves the main insights unchanged. The main technical difference is that the marginal cost now depends on w and r, and that the labor market clearing condition must be adjusted. The first step is to specify the monitoring technology. I assume a monitoring technology with constant returns to scale:

Homogenous Intermediation Hypothesis (HIH). *Monitoring an entrepreneur of type μ with capital A requires μA units of intermediation. Intermediation is produced with the constant return technology $F_\phi(A\bar{n}_\phi, K_\phi)$ with $\bar{n}_\phi$ units of labor and K_ϕ units of capital.*

Abstracting from capital, HIH simply says that it takes a given amount of human capital in banking to monitor one unit of entrepreneurial human capital.[15] The following proposition summarizes the properties of the equilibrium.

Proposition 3. *Equilibrium Corporate Finance*. *Under homogenous intermediation, the income share of finance is constant on the Balanced Growth Path, and depends on the growth rate of the economy only through the equilibrium real rate. The consequences of shifts in supply and demand for intermediation are as described in Proposition 2.*

[14] Students of trade theory might naturally wonder if there is an analogy between financial friction and shipping costs. Estevadeordal, Frantz, and Taylor (2003) show that real transport costs rose during the inter-war period. In their framework, let A be productivity, so the real cost of production is $1/A$, and let τ be the real of cost transportation. The price of domestic goods is $1/A$. The price of imported goods is $1/A + \tau$. The relative price of imported goods is $1 + \tau A$. If there is productivity growth in production but not in transportation, the relative price of imports goes up. Is there an analogy with finance? In the corporate finance context, one could think of using internally generated cash as buying domestic goods, and raising external capital as buying a foreign good. However, the shipping cost result is based on a very particular technology. It assumes that the physical characteristics of the goods are constant. Say a car weighs one ton, and making a car requires $1/A$ units (of numeraire). When we say that the transport cost is τ we mean that it takes τ units to ship one ton. Productivity growth does not decrease the weight of the car and does not create more value (safety, speed) per ton. It simply allows more cars to be built by fewer people. But in finance "shipping" one million or one billion costs the same. The issue is the monitoring technology. The most natural monitoring technology says that it takes one banker to monitor $1/\psi$ borrowers. Each borrower has a productivity A. Without monitoring they cannot borrow the optimal amount and their output is only $(1-\chi)A$ (which might be zero if $\chi = 1$). Borrowers are willing to pay up to χA for financial services. The costs of services is ψW the wage of the banker times the monitoring requirement. Demand for services therefore depends on $\frac{\chi A}{\psi W}$. On the balanced growth path wages grow with A so the finance share is constant and independent of technology in the non financial sector. This intuition is simple: the value of financial services is proportional to the productivity of the agents who receive these services. In other words, for a given monitoring technology, the effective productivity of financial intermediation is automatically indexed on the productivity of the non financial sector. This digression was inspired by Paul Krugman who was my discussant at the 2011 New York Area Monetary Policy Workshop.

[15] A comparison of the respective careers of Charles Ponzi and Bernard Madoff suggests that fraud and detection have remained similar.

Proof. See Appendix. □

In Section 5 I will test HIH and show that it is supported by the data. Finally, it is straightforward to extend the model to the case where entrepreneurs have cash-on-hand so that not all external finance is less than capital. Such a model is presented in the Appendix and used for quality adjustments in Section 5. For future references, I define $b_k \leq k$ as external finance.

3 Household Finance: Money and Credit

A significant share of the recent increase in credit markets is due to household borrowing. In addition, the finance industry provides liquidity and payment services to households (and savers more generally). We therefore need to extend the model to take into account household debt and liquidity. In practice, much of household debt has an important life-cycle component (i.e., mortgages), so we need to extend the model to overlapping generations. I consider a mixed model with two types of households: one type is infinitely lived, the other lives for two periods.[16] The first type of household is the neoclassical household depicted in the first section. The second type of household is part of an OLG structure where agents live for two periods. Households in the model do not lend directly to one another. Savers lend to intermediaries, and intermediaries lend to firms and households.

Liquidity Services and Long-Lived Household

This long-lived household (index 0) owns the capital stock but has no labor endowment. Liquidity services can be modeled using a cash-in-advance framework (see Lucas (1980), Lucas and Stokey (1987), and Sargent and Smith (2009) for instance), or with money in the utility function. I use the later for simplicity. The flow utility is now $u\left(C_t, M_t\right) = \frac{\left(C_t M_t^{\nu}\right)^{1-\rho}}{1-\rho}$. The budget constraint becomes $S_{t+1} + C_t + \psi_m M_t \leq \left(1 + r_t\right) S_t$, where ψ_m is the price of liquidity services, and S are total savings.[17] The Euler equation $u_C\left(t\right) = \beta \mathbb{E}_t\left[\left(1 + r_{t+1}\right) u_C\left(t+1\right)\right]$ becomes

$$M_t^{\nu(1-\rho)} C_{0,t}^{-\rho} = \beta \mathbb{E}_t\left[\left(1 + r_{t+1}\right) M_{t+1}^{\nu(1-\rho)} C_{0,t+1}^{-\rho}\right].$$

The liquidity demand equation $u_M\left(t\right) = \psi_m u_C\left(t\right)$ is simply

$$\psi_m M_t = \nu C_{0,t}.$$

[16] Several reasons motivate this choice. First, the simplest OLG model with two periods is not appealing because households do not actually borrow: the young ones save, and the old ones eat their savings. Second, bequests are of first order importance empirically. The simplest way to capture all these ideas is the mixed model. The interpretation is that the long-lived household has perfect bequest motives, so it is equivalent to an infinitely lived agent. Mehra, Piguillem, and Prescott (2011) study a model where household save for retirement over an uncertain lifetime.

[17] Liquidity services enter essentially as another good consumed by the household. Note that the household saves S for a gross return of $1 + r$. Deposits yield $\left(1 + r\right) / \left(1 + \psi_m\right)$. So this model assumes a constant spread between the rate on deposits and the rate on "illiquid" savings.

On the BGP, M grows at the same rate as C. The Euler equation becomes $1 = \beta \mathbb{E}_t \left[(1 + r_{t+1}) \left(\frac{C_{t+1}}{C_t} \right)^{\nu(1-\rho)-\rho} \right]$, so the equilibrium interest rate solves

$$\beta (1 + r) = (1 + \gamma)^\theta ,$$

where $\theta \equiv \rho - \nu (1 - \rho)$.

Household Credit

The other households live for two periods. The young (index 1) have a labor endowment η_1 and the old (index 2) have a labor endowment η_2. We normalize the labor supply to one: $\eta_1 + \eta_2 = 1$. The life-time utility of a young household is $u(C_{1,t}, M_{1,t}) + \beta u(C_{2,t+1}, M_{2,t+1})$. I consider the case where they want to borrow when they are young (i.e., η_1 is small enough). In the first period, its budget constraint is $C_{1t} + \psi_m M_{1t} = \eta_1 W_{1t} + (1 - \psi_c) B_t$. The screening costs of lending to households is ψ_c. In the second period, it consumes $C_{2t+1} + \psi_m M_{2t+1} = \eta_2 W_{t+1} - (1 + r_{t+1}) B_t$.

The detrended Euler equation for short-lived households is

$$\mathbb{E}_t \left[\beta (1 + r_{t+1}) (1 + \gamma_{t+1})^{-\theta} \left(\frac{c_{2t+1}}{c_{1t}} \right)^{-\theta} \right] = 1 - \psi_c.$$

Since, on the balanced growth path, the interest rate is pinned down by the long horizon savers at $\beta (1 + r) = (1 + \gamma)^\theta$, the Euler equation of short lived households becomes simply

$$c_1 = (1 - \psi_c)^{\frac{1}{\theta}} c_2.$$

In addition, we have $\psi_m m = \nu c$ for each cohort. The budget constraints are therefore $(1 + \nu) c_1 = \eta_1 w + (1 - \psi_c) b$ and $(1 + \nu) c_2 = \eta_2 w - \frac{1+r}{1+\gamma} b$.

Equilibrium

For simplicity I focus here on the case of homogenous corporate borrowers, Cobb-Douglass technology, and complete intermediation of capital ($b_k = k$). The capital demand equation is $(1 - \alpha) \frac{y_t}{k_t} = r_t + \delta + \psi_k$. Capital is still given by $(1 - \alpha) k^{-\alpha} = r + \delta + \psi_k$, and the capital-output ratio: $\frac{k}{y} = \frac{1-\alpha}{r+\delta+\psi_k}$.

Using $w = \alpha y$, we can use the Euler equations and budget constraints to compute the borrowing by young households relative to GDP:

$$\frac{b}{y} = \frac{(1 - \psi_c)^{\frac{1}{\theta}} \eta_2 - \eta_1}{1 - \psi_c + (1 - \psi_c)^{\frac{1}{\theta}} \frac{1+r}{1+\gamma}} \alpha.$$

If ψ_c is 0, we have $c_1 = c_2$, and consumption growth is the same for all agents and equal to γ, the fundamental growth rate of the economy. From the perspective of current consumption, borrowing costs act as a tax on future labor income. If ψ_c is too high, no borrowing takes place and the consumer credit market collapses.

To close the model, we obtain the consumption of the long lived savers (c_0) from the resource constraint. Since $\psi_m m = \nu c$ for all agents, we have $y = (1+\nu)(c_0 + c_1 + c_2) + \psi_c b + (\gamma + \delta + \psi_k) k$. Since $\eta_1 + \eta_2 = 1$, and $w = \alpha y$, we get

$$(1+\nu) c_0 = (r - \gamma) \left(k + \frac{b}{1+\gamma} \right).$$

Total expenditure of long-lived households is equal to their capital income from loans to corporates and to short-lived households.

The GDP share of finance is

$$\phi = \psi_m \frac{m}{y} + \psi_c \frac{b_c}{y} + \psi_k \frac{b_k}{y}.$$

We summarize our results in the following Proposition.

Proposition 4. ***Equilibrium Intermediation with Corporate and Household Finance***. *Under HIH, the income share of finance is constant on the Balanced Growth Path. Consumer borrowing over GDP (b/y) decreases with intermediation costs ψ_c and increases with the slope of life-cycle earnings profiles (η_2/η_1). There is no permanent crowding out of corporate investment by household borrowing. Liquidity demand (m/y) decreases with ψ_m and ψ_k. Changes in ψ_m have no impact on the GDP share of finance, while changes in ψ_c have an inverse U-shape impact.*

The no-crowding-out result holds on the Balanced Growth Path. It relies on a constant real rate and the implied portfolio adjustments by long-lived agents. The bigger the ratio η_2/η_1 the larger the borrowing. For instance, increased years schooling generates more borrowing, and a larger financial sector. Improvements in corporate finance increase liquidity demand because they increase the consumption output ratio. When ψ_k goes down, k/y goes up while b/y is unchanged, therefore $\nu c_0/y$ goes up. One important point is that, under HIH, the model does not predict an income effect, i.e., just because a country becomes richer does not mean that it should spend a higher fraction of its income on financial services. I will test directly HIH in Section 5.

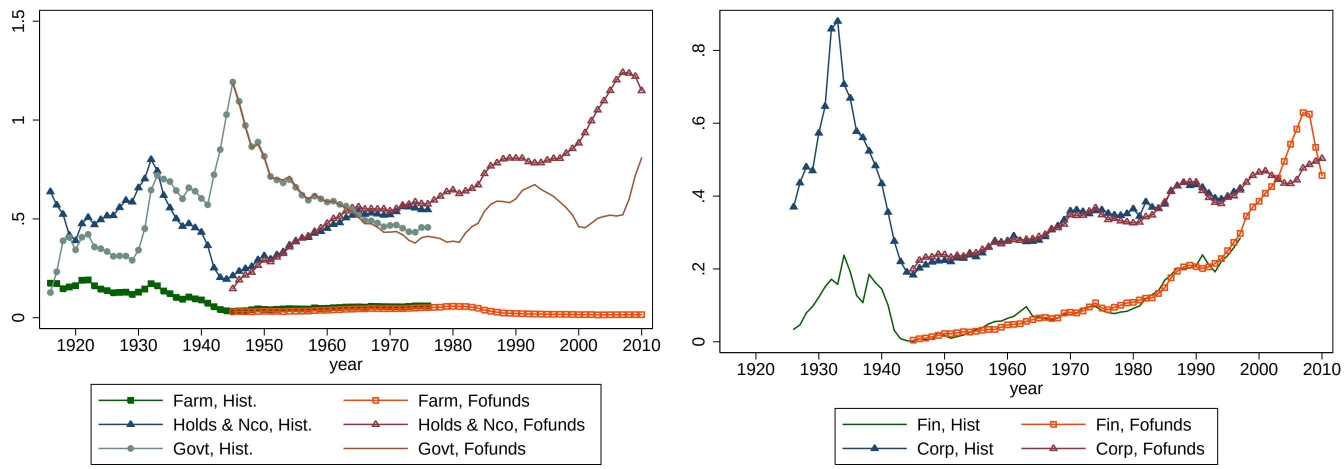
4 Output Series

The previous sections have made it clear that we need to carefully consider the production of assets, and take heterogeneity into account. In this section I construct empirical proxies for $\frac{m}{y}; \frac{b_c}{y}; \frac{b_k}{y}$.

4.1 Credit Markets

Figure 4 presents credit liabilities of farms, households and the government. The first point to take away is the good match between the various sources. As with the income share above, this allows us to extend the series in the past. The main features of the long run series in Figure 4 are the impact of WWII on government debt, and the growth of household debt in the post-war era. The bottom panel of Figure 4 presents credit liabilities of financial and non-financial corporates. Two features stand out. First, the non-financial corporate credit market is not as deep even today as it was in the 1920s. Second, financial firms have recently become major issuers of debt. Banks used to fund themselves with deposits and equity, and almost no long term debt. Today they issue a lot of long term debt. Note that it is critical to separate financial and non-financial issuers. What should count as output for the finance industry are only issuances by non-financial firms.

Figure 4: Debt over GDP

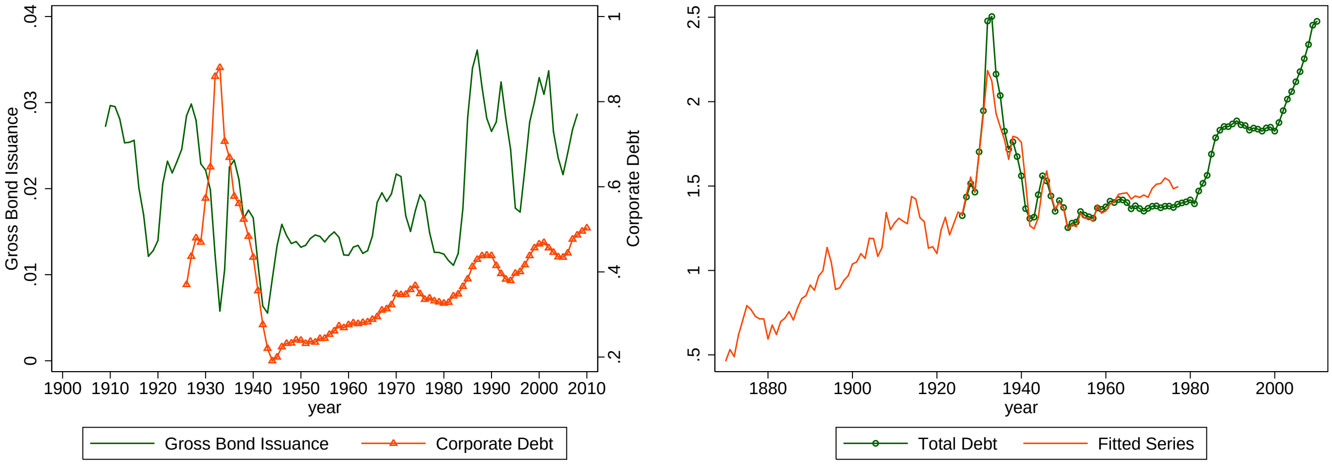


Notes: Farm is simply the farm sector, "Holds & Nco" include households and non-corporate, "govt" is government. "Fin" is finance and real estate, "corp" is non-financial corporate. "Fofunds" is flow of funds, "Hist" is Historical Statistics of the United States.

In the theory outlined earlier, there is no distinction between outstanding assets and new issuances. In the data the two can be different and it is useful to consider stocks and flows separately. The left panel of Figure 5 shows the issuance of (non-financial) corporate bonds and the ratio of outstanding (non-financial) corporate debt over GDP. Corporate debt includes loans and other liabilities (mortgages, etc.) so it corresponds to a larger stock than the one implied by bond issuances alone (bonds are typically about half of credit market instruments issued by the corporate sector). There is a clear post-war correlation, but also a significant difference in the 1930s when the debt to GDP ratio peaks, in part because of deflation, while bond issuance collapses.

To extend the credit series before 1920, I use the balance sheets of financial firms. I measure assets on the balance sheets of commercial banks, mutual banks, savings and loans, federal reserve banks, brokers, and life insurance companies. I define total assets as the sum of assets of all these financial firms over GDP. I use this series to extend the total non-financial debt series (households & non corporates, farms, corporates, government). I regress total credit on total assets and use the predicted value to extend the credit series. The fit and the extended series are presented on the right panel of Figure 5.

Figure 5: Non-Financial Debt



Notes: "Gross Issuance" is a three-year centered moving average of gross bond issuance by non-financial firms. Sources are Historical Statistics of the United States and Baker and Wurgler (2000). Fitted Series uses assets on balance sheets of financial firms to predict total debt. Sources are Historical Statistics of the United States and Flow of Funds.

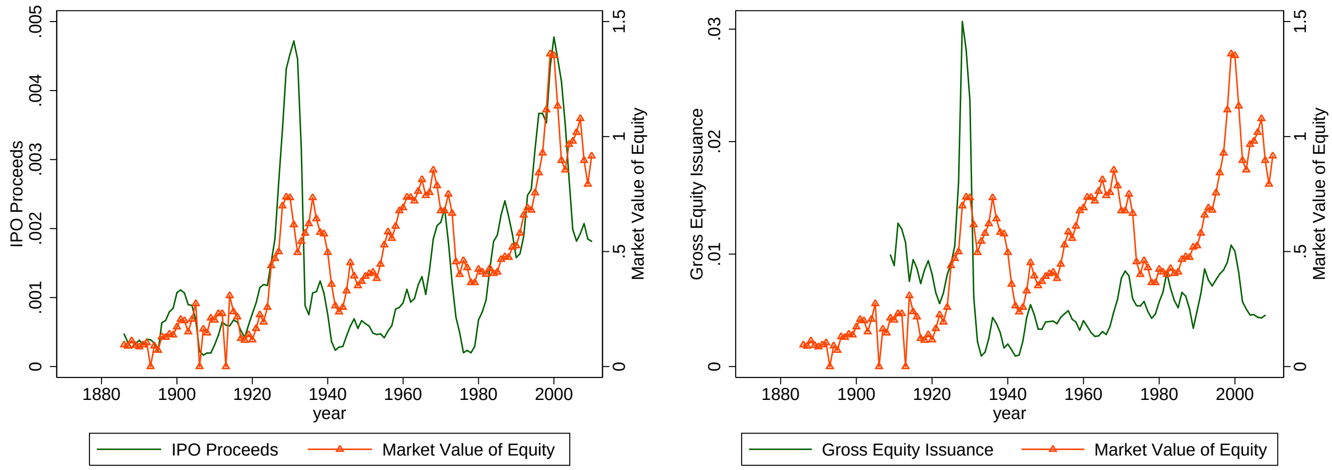
4.2 Equity Market

The equity market is difficult to deal with because of valuation effects. Stocks, unlike bonds, are recorded at market value. The ratio of the market value of equity to GDP can fluctuate without any intermediation services (i.e., without any issuance of equity). Another problem is that net issuances are often negative. However, negative net issuances do not imply that no intermediation services are produced. To deal with these problems I use three measures of equity production: total market value over GDP, IPO proceeds over GDP, and gross (non-financial) equity offerings over GDP. The advantage of using IPOs is that they provide a good measure of entry and growth by young firms, whose screening and monitoring requirements are certainly higher than those of established companies. Thus, the IPO series will allow me to control for heterogeneity as required by the theory of Section 2. Heterogeneity adjustment are presented in the next section.

Figures 6 shows the IPO and market value measures. As argued by Jovanovic and Rousseau (2001), the IPO market of the 1920s was remarkably active, even compared to the one of the 1990s: the IPO firms are of similar ages, and the proceeds are comparable. By contrast, the market value of GDP has an upward trend, in part due to increases in price earnings ratios, and in part due to an increase in the stock market listing.

The right panel of Figure 6 shows the evolution of the market value of equity and of gross issuances of stock. The gross issuance series is dominated by the large spike in the late 1920s, but even without this spike, the level of stock offerings appears remarkably stationary. Thus, contrary to common wisdom, an increase in equity market value does not necessarily imply that equity funding becomes more important.

Figure 6: Equity Value, IPO, and Gross Issuance over GDP



Notes: Market Value of non-financial corporate firms from the Flow of Funds and from CRSP. IPO is a three-year centered moving average of IPO proceeds over GDP. Sources are Jovanovic and Rousseau (2001) and Ritter (2011). Gross issuance is a three-year centered moving averages. Sources are Historical Statistics of the United States and Baker and Wurgler (2000).

4.3 Money and Liquidity

In addition to credit (on the asset side of banks), households, firms and local governments benefit from payment and liquidity services (on the liability side of banks and money market funds). For households, I use the total currency and deposits, including money market fund shares, held by households and nonprofit organizations. The left panel of Figure 7 shows the evolution of this variable.

An important issue in the measurement of liquidity provision is the rise of the shadow banking system. Gorton, Lewellen, and Metrick (2012) argue that a significant share recent activities in the financial sector was aimed at creating risk free assets with money-like features. For firms (incorporated or not), I follow Gorton, Lewellen, and Metrick (2012) and I treat repos as shadow deposits. The series is thus the sum of checkable deposits and currency, time and savings deposits, money markets mutual funds shares, and repos (by non financial firms).[18]

4.4 Aggregation

The models presented in Sections 2 and 3 suggest to measure the output of the finance industry as the weighted sum of assets intermediated. I normalize the weight on consumer credit to one. Finance output is then $y^{\phi} = b_c + \frac{\psi_k}{\psi_c} b_k + \frac{\psi_m}{\psi_c} m + \frac{\psi_g}{\psi_c} b_g$, where b_k is corporate intermediation, b_c is household borrowing, m are deposits, and b_g is government debt. An important issue is that I cannot measure the GDP share linked to these assets separately, so I cannot directly estimate the parameters ψ's.[19] I will therefore first assume that the relatives ψ's are constant,

[18] I have experimented with an adjustment for the fact that deposit insurance provided by the government makes it cheaper for private agents to create deposits. The adjustments seem rather arbitrary and did not make a significant difference so I dropped it. But more quantitative work would clearly be needed here.

[19] This is not a simple accounting issue. Financial tasks are deeply intertwined. Insurance companies and pension funds perform their own independent credit analysis. Banks act as market makers. Investment banks behave as hedge funds. In addition, the mapping from industry to tasks has changed over time. XX

Figure 7: Deposits and Shadow Deposits

Notes: Sources are Historical Statistics of the United States and Flow of Funds.

while allowing the average ψ so move over time. I will later provide quality adjustment that change the relative ψ's.

Corporate intermediation itself is made of equity and debt financing, hence $\psi_k b_k = \psi_k^d d_k + \psi_k^e e_k$, where e_k is corporate equity and d_k is corporate debt. I construct two series for each type of output, one of flows of new assets, and one of levels of outstanding assets. Before doing so, I need to discuss M&A activities.

M&As

An important activity of financial intermediaries is advising on mergers and acquisitions. Rhodes-Kropf and Robinson (2008) show that M&As differ from other types of investment and require specific search efforts. From 1980 to 2010, I use data from SDC and Bloomberg to compute the value of merger deals. I then use historical data from Jovanovic and Rousseau (2005) to extend the series back to 1890. The next step is to apply the proper weight to the M&A series. I use these to construct my M&A series. To scale it, I assume that merger fees are 2% of the volume. The industry standard range for 1% for large deals to 4% for smaller ones. My assumption of 2% is probably a bit too high, but there are also probably some ancillary activities associated with mergers.

Level Measure

As a benchmark, I will assume that $\psi_k^d = \psi_c$ so that it is equally difficult to extend credit to firms or to households. I will also assume $\psi_m = \psi_c$ so that the spreads on both sides on banks balance sheets are similar. There is also the issue of the debt of the government. On the one hand, it is risk-free and liquid, and it might actually help the functioning of financial markets (Krishnamurthy and Vissing-Jorgensen (2010), Greenwood, Hanson, and Stein (2011)). So one option would be to ignore government debt, i.e., to assume that it does not require intermediation services. On the other hand, there is some duration risk, and it needs to be traded. Somewhat arbitrarily, I assume that government debt requirements are 1/10 of that of private debt: $\frac{\psi_g}{\psi_c} = 0.1$.

$$y^{\phi}_{level} = b^{level}_c + d^{level}_k + e^{level}_k + m^{level} + 0.1 b^{level}_g + y^{level}_{M\&A}$$

Flow Measure

Flows correspond to issuances. I calibrate the relative weights using underwriter fees.Altinkilic and Hansen (2000) report fees of 3% to 4% for equity and about 1% for bonds. For flows, I therefore assume $\psi^e_k = 3.5\psi^d_k$. These numbers are also in line with recent reports by large investment banks (see discussion of JP Morgan in Appendix). Since I do not have a separate measure of flows for deposits, I use the scaling factor $\lambda = \bar{y}^{\phi}_{flow}/\bar{y}^{\phi}_{level}$ $m^{flow} = \lambda m^{level}$. Similarly, since I do not have a level measure for M&As, I use $y^{level}_{M\&A} = y^{flow}_{M\&A}/\lambda$.

$$y^{\phi}_{flow} = b^{flow}_c + d^{flow}_k + 3.5 e^{flow}_k + m^{flow} + 0.1 b^{flow}_g + y^{flow}_{M\&A}$$

Composite measures

As explained above, I construct two output series for the finance industry. One using the flows (gross issuances over GDP) and one using the levels (debt and equity over GDP). Note that both are relevant in theory. Screening models apply to the flow of new issuances, while monitoring models apply to the stocks. Trading applies to both. The two series are displayed in the left panel of Figure 8. The production of financial services increases steadily until WWI, and rapidly after 1919 until 1929. It collapses during the great depression and WWII. It increases steadily until 1975 and more randomly afterwards. The flow and level measures share the same trends, but there are clear differences at medium frequencies. The flow measure is more stationary that the level measure. The flow measure collapses quickly during the great depression while the level measure peaks later in 1932-33 (this is exacerbated by deflation). A similar pattern seems to emerge during and after the great recession of 2008-2009.

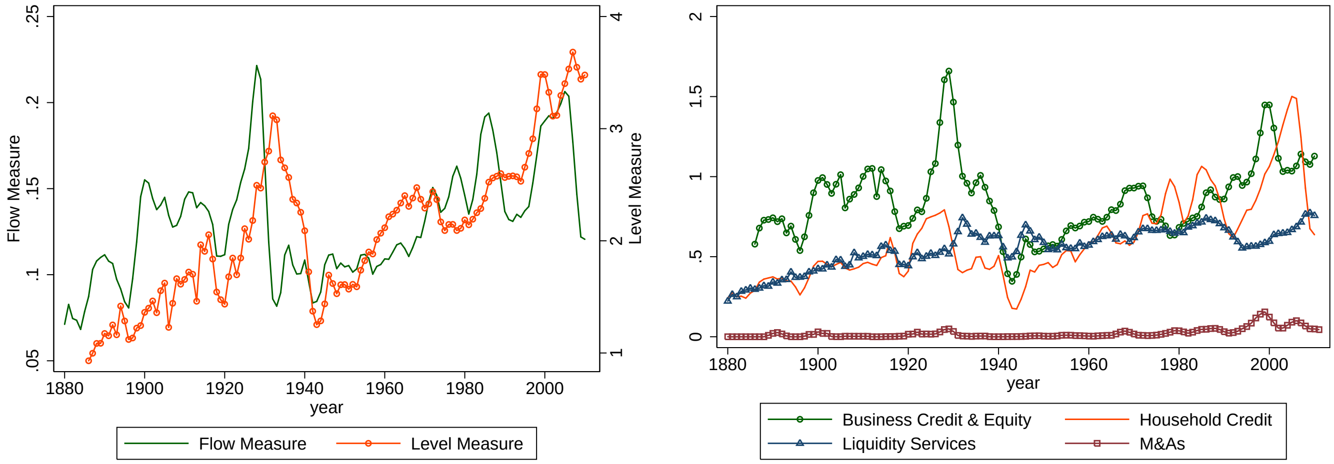
I now wish to combine the flow and level measures. Since there is no reason to prefer one to the other, it seems logical to put equal weights on the two measures. On average, in the post-1950 period (where the data is most reliable), the ratio of the flow measure to the level measure is 0.062. In other words, the level measure is about 16 times the flow one. I therefore set $\lambda = \bar{y}^{\phi}_{flow}/\bar{y}^{\phi}_{level} = 1/16$ and I construct the composite measure as

$$y^{\phi} = \frac{1}{2}\left(y^{\phi}_{level} + \frac{y^{\phi}_{flow}}{\lambda} \right) \tag{9}$$

Note that, for convenience, the scale the composite measure is comparable to the scale of the level measures. The unit cost of intermediation can then naturally be compared to an interest rate spread. Changing the weight changes the short run behavior of the composite measure, but not its long run behavior.

The right panel of Figure 8 presents the composite measures corresponding to 4 broad functions discussed earlier:

Figure 8: Financial Intermediation Output



Notes. Left: aggregate measures of output for US finance industry, levels and flows, as shares of GDP. Right: composite measures (average of stock and flows) across broad functions.

credit and equity intermediation services to firms (farms, corporate, non-corporate), credit intermediation services to households, liquidity services to both, and M&A activities. In each case, the composite measure is based on the stocks and flow measures aggregated as in (9). Note that the liquidity and M&A measures are not in fact composite measures since I do not have independent flow measures for deposits or level measures for M&As. It is clear from this figure that credit intermediation for firms and households are the most volatile series. There is also a significant increase in liquidity services in the 2000s. M&As play a significant role mostly in the 1990s.

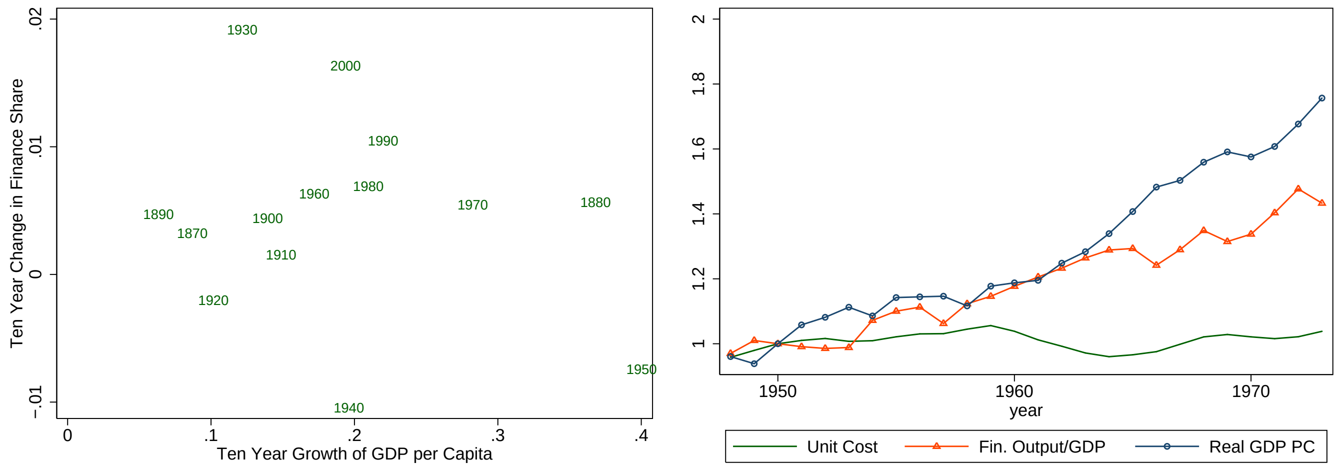
5 The Unit Cost of Intermediation

5.1 Evidence of Constant Returns to Scale

As shown earlier, under HIH, the theory does not predict that the finance share of GDP should mechanically increase with income. The left panel of Figure 9 shows that this lack of mechanical relation is supporter by the evidence. Of course, income per capita is not a stationary variable, while the share of finance is bounded by construction. We can nonetheless look at long run changes in income and ask whether the share of finance tends to increase more when income grows more quickly. The left panel of Figure 9 shows that there is no relationship between the growth of income per capita and the share of finance. Similarly, Bickenbach, Bode, Dohse, Hanley, and Schweickert (2009) show that the income share of finance has remained remarkably constant in Germany over the past 30 years. More precisely, using KLEMS for Europe (see O'Mahony and Timmer (2009)) one can see that the finance share in Germany was 4.3% in 1980, 4.68% in 1990, 4.19% in 2000, and 4.47% in 2006.

Looking at the (lack of) correlation between income per-capita and the share of finance is a logical first test because it is simple and does not require any measure of finance output. But it is limited and does not provide direct evidence for constant returns to scale. I now present a direct test. The right panel of figure 9 presents evidence consistent with HIH. It uses the period 1947-1973, for two reasons. First, the post-war data is reliable. Second, as I will discuss shortly, quality adjustments are not important over this period. Since these adjustments are difficult to implement, it is more convincing to present the evidence without them.

Figure 9: No Income Effect and Constant Returns to Scale



Notes: Unit cost of financial intermediation, composite measure of finance output, and real GDP per capita. Series normalized to one in 1950.

From 1950 to 1980, real GDP per-capita increases by 80% and the ratio of finance output to GDP increase by more than 50%, but my estimate of the unit cost of intermediation remains roughly constant (all series are presented as ratios of their 1950 values). In other words, people are a lot richer, financial markets are a lot larger, but the unit cost does not change. This provides fairly strong support for HIH.

5.2 Quality Adjustments

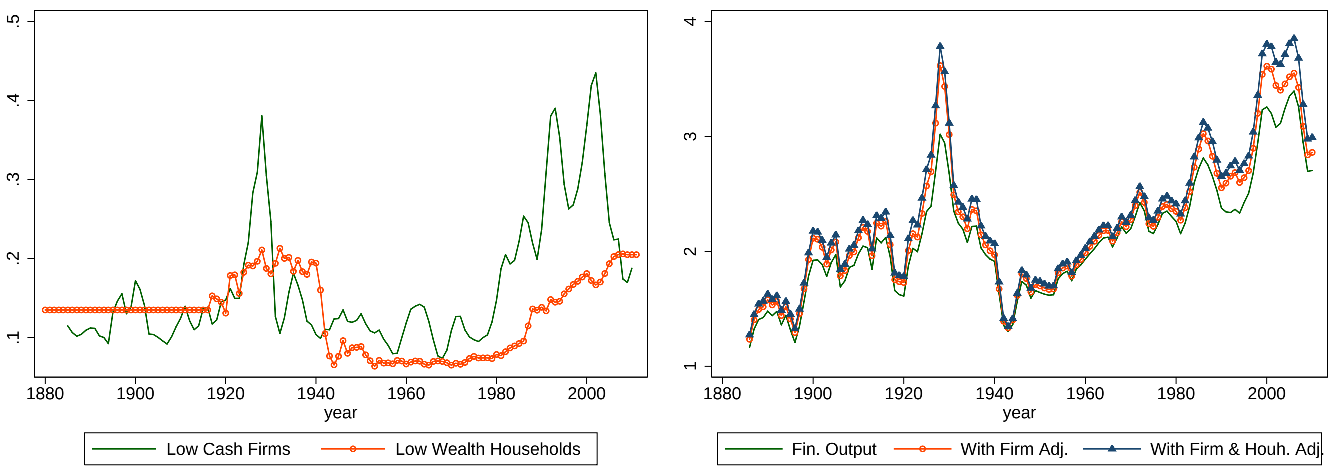
The theory of Sections 2 and 3 makes it clear that the quantities of intermediation should be adjusted for the difficulty of monitoring/screening borrowers. To be concrete, suppose that borrowers become harder to screen. Then a given amount of lending requires more intermediation. As a result, the income share of finance will increase. Without quality adjustment, my measure would register a spurious increase in intermediation cost (i.e., a spurious decrease in financial efficiency). I consider two separate adjustments, one for corporate borrowing, and one for household borrowing. The two adjustments are displayed on Figure 10.

For corporate borrowing, the appendix provides a simple adjustment based on the aggregate importance of cash-poor firms. Philippon (2008) uses Compustat to construct a empirical proxy for this share, namely the share of aggregate investment that is done by firm who must borrow more than 3/4 of their capital spending. This measure is thus available from 1950 to 2010. To extend the series in the past, I use IPO proceeds. In the post-war period, when both series are available, they are highly correlated. Once this is done, I use a calibrated version of the corporate finance model to compute the required adjustment. The model suggests an adjustment of 1.25 times the share of low cash firms, with a normalization so that aggregate monitoring equals the composite measure of firm intermediation when the share is zero. If the share of low cash firms is 20%, the model says that the true production of intermediation services is 25% higher than the composite credit measure. The adjustment is therefore

$$\hat{b}_k = (1 + 1.25 \times low_k) \times b_k$$

where *low* is the share of low cash firms and b is the unadjusted measure of firm credit.

Figure 10: Quality Adjustments



Notes: Sources are Historical Statistics of the United States and Flow of Funds.

A similar issue arises with households. First, on a per-dollar basis, it is more expensive to lend to poor households than to wealthy households. Second, relatively poor households have gained access to credit. Using the Survey of Consumer Finances, Moore and Palumbo (2010) document that from 1989 to 2007 the fraction of households with positive debt balances increased from 72% to 77%. This increased was concentrated in the bottom of the income distribution. For households in the 0-40 percentiles of income, the fraction with some debt outstanding goes from 53% to 61% between 1989 and 2007. In 2005 Mayer and Pence (2008) show that subprime originations were between 15% and 20% of all HMDA originations. The challenge is twofold. First, one must estimate the share of low wealth households over the sample period. Second, I must translate this share into a monitoring requirements. Similar calculations to the one for firms suggest an adjustment of roughly 1:

$$\hat{b}_c = (1 + 1 \times low_c) \times b_c$$

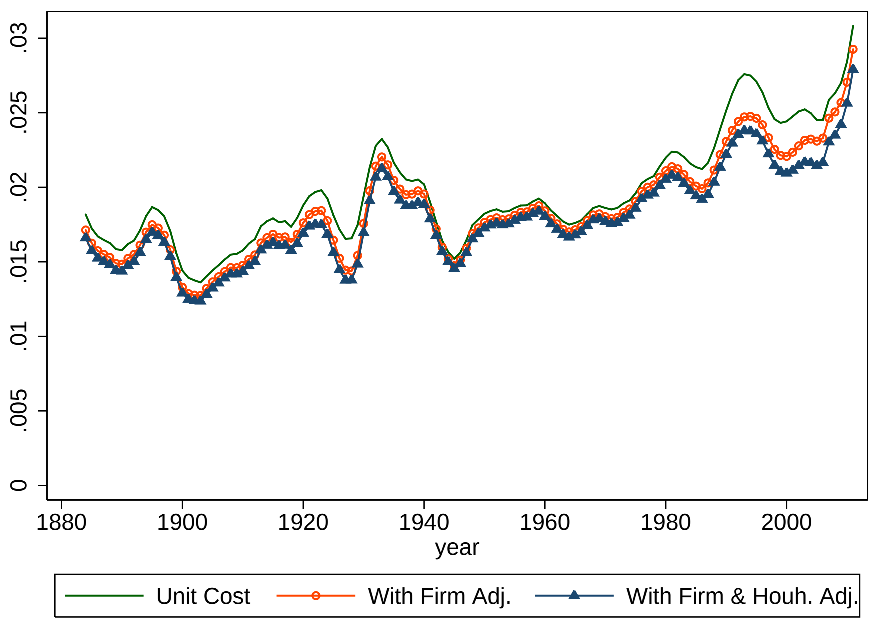
To extend the series I assume that the share of low wealth household is proportional to inequality, and I use data from Piketty and Saez (2003) to extend the series.

5.3 Unit Cost

Figure 11 estimates the cost of financial intermediation, defined as income divided by output. For output, I use the composite measure of equation (9). Figure 11 is the main contribution of the paper. It brings together the theory of Sections 2 and 3 and the historical/empirical work of Section 4. There are two main points to take away. The first is that the ratio is remarkably stable. Recall that we started from a series in Figure 1 that fluctuates by a factor of 5 (9% relative to less than 2%). All the debt, deposit and equity series also vary a lot over time. But their ratio, properly scaled, seems quite stable. On Figure 11 it stays roughly between 1.5% and 2.5% over 130 years. One must also keep in mind that the model ignores business cycles and assumes constant real interest rates.

The second main point is that the finance cost index has been trending upward, especially since the late 1970s. This is counter-intuitive. If anything, the technological development of the past 40 years (IT in particular) should have disproportionately increased efficiency in the finance industry. How is it possible for today's finance industry not to be significantly more efficient that the finance industry of John Pierpont Morgan? I conclude that Figure 11 presents a puzzle for future research. In the next (and last) section, I discuss some recent or ongoing research that might shed light on this puzzle.

Figure 11: Financial Intermediation Cost Index



Notes: Production of credit assets, equity and deposits divided by total intermediation costs.

6 Discussion

6.1 Information Technology

An obvious driving force in financial intermediation is information technology. The models of Sections 2 and 3 predict that technological improvement should typically lower the share of GDP spent on financial intermediation. In particular, this prediction is unambiguous for intermediation services used by all firms. The reason is that for these basic services there is no extensive margin effect where better finance could gives access to firms that were previously priced out. This seems like a fair description of at least some part of retail finance. Essentially, the argument is that IT must lower the physical transaction costs of buying, pooling and holding financial assets. An apt analogy is with retail and wholesale trade. Indeed, Philippon (2012) shows that sharply increasing IT investment in wholesale and retail trade has coincided with lower (nominal) GDP shares for these sectors, exactly as theory would predict. Exactly the opposite is true in finance. IT therefore seems to deepen the puzzle instead of solving it, but more work is clearly needed. Perhaps it is the impact of IT on trading that really matters.

6.2 Price Informativeness and Risk Sharing

Using the GDP share of finance to measure the costs of financial intermediation is fairly straightforward. It ignores hidden costs of systemic risk, but it captures all fees and spreads. The output measure developed above, however, might not fully capture the production of information via prices, and the provision of insurance. Going back to the models of Section 2 and 3, it is important to ask the following question: If improvements in financial intermediation lead to more informative prices or better risk sharing, where would these improvements be seen in equilibrium?

Informativeness of Prices

The simplest way to test the hypothesis that prices have become more informative is to directly test the signal-to-noise ratio of asset prices. Bai, Philippon, and Savov (2011) ask if current firm-level equity and bond prices predict future firm productivity and if this forecasting power has changed over time. Preliminary evidence suggest that the forecasting power has been remarkably stable over the past 50 years (for comparable sets of firms). In other words, while bid-ask spreads have decreased, and while many have claimed that financial markets have become more liquid, this does not appear to have translated into “better” prices. I am not aware of direct evidence regarding other classes of asset prices. For commodity prices, some practitioners (e.g., Hadas (2011)) seem to argue that prices have become less informative. See also Tang and Xiong (2011) for recent research on commodity prices.

Risk Management and Derivatives

Another benefit of financial intermediation is risk sharing. Risk sharing can affect firms and households. At the firm level, risk sharing is commonly called risk management. Better risk management would, in equilibrium, mostly translate into lower cost of fund, more issuances and more investment. This first effect would be captured by our measures of debt and equity issuances. Better risk management could also increase TFP if high productivity projects are also riskier.

The market for financial derivatives is extremely large. Since these contracts are in zero net supply, however, they do not enter directly into the calculation of output for the finance industry. How should we account for these contracts? One thing is clear: it would not make sense to count derivatives at face value.[20] In terms of economic theory, derivatives can add real value in one of two ways: (i) risk sharing; (ii) price discovery. Risk sharing among intermediaries would not create a bias in my measurements, however. It would simply lead to lower borrowing costs and cheaper financing.[21] Therefore, the only bias from derivative contracts must come from better risk sharing or price discovery among non-financial borrowers.

The correct way to measure the value added of derivatives is to measure directly the informativeness of prices, or the welfare gains from risk sharing among non-financial firms and households. I have already argued that there is no evidence of improvement in informativeness of prices. I am also not aware of any direct evidence suggesting significant improvements in corporate risk management. One obvious index, that of precautionary savings by businesses, suggest even the opposite: corporate cash holdings have increased over the past 30 years. There is also no direct evidence of credit derivatives leading to better risk management, and it is commonly believed that hedging represents only a small fraction of all trades in the CDS market.[22]

Risk Sharing & Consumption Smoothing

At the household level, better risk sharing should lead to less consumption risk. Income inequality has increased dramatically in the U.S. over the past 30 years. If financial markets have improved risk sharing, however, one would expect consumption inequality to have increased by less than income inequality. This is a controversial issue, but Aguiar and Bils (2011) find that consumption inequality has closely tracked income inequality over the period 1980-2007. It seems difficult to argue that risk sharing among households has improved significantly over time. It is also difficult to point to a financial innovation in the past 30 years that would have directly improved risk sharing opportunities among households.

There is evidence of improved consumption smoothing in the housing market. Gerardi, Rosen, and Willen (2010) find that the purchase price of a household's home predicts its future income. The link is stronger after 1985, which

[20] For instance, consider the following example. Without derivatives, corporation A borrows from bank B and bank B retains the credit and duration risks on its books. With derivatives, bank B buys insurance against credit risk from fund C using a CDS. The sum of B and C holds exactly the same risk. Absent other frictions, the two models are exactly equivalent.

[21] To see why let us go back to the simple example. Suppose there are frictions that rationalize why B and C should be separate entities, and why they gain from trading with each other (i.e., B has a comparative advantage at managing duration risk, and C at managing credit risk). Then the existence of CDS contracts can improve risk sharing among intermediaries, lower the risk premia, and lead to a decrease in the borrowing costs of A. Hence, with free entry, the total income going to intermediaries B+C would decrease. This could then increase the demand for borrowing, as explained in Section 2. All these effects would be captured by the model: either borrowing costs would go down, or borrowing volumes would go up, or both. In all cases, my approach would register an increase in efficiency.

[22] Also remember that the liquidity measure captures deposits and repos contracts by non financial firms. So cash management is accounted for by my measures.

coincides with important innovations in the mortgage market. The increase in the relationship is more pronounced for households more likely to be credit constrained. This type of smoothing is captured by the model because I measure all mortgage borrowing. So unlike pure insurance, consumption smoothing over the life cycle does not create a bias in my estimation.

6.3 Trading

At this point, we are left with a puzzle. Finance has obviously benefited from the IT revolution and this has certainly lowered the cost of retail finance. Yet, even accounting for all the financial assets created in the U.S., the cost of intermediation appears to have increased. So why is the non-financial sector still transferring so much income to the financial sector?

One proximate cause might be trading. Trading costs have decreased (Hasbrouck (2009)), but trading volumes have increased even more. In addition, the costs of active fund management are large. French (2008) estimates that investors spend 0.67% of asset value trying (in vain, by definition) to beat the market. French's calculations are only for the equity market. In Figure 11, a drop in the intermediation cost index of 50 to 60 basis points would indeed bring it back towards its historical average. With output at 4 times GDP, this suggests that about 2% of GDP, or about $280 billions annually, are either wasted or at least difficult to account for.

Why do people trade so much? Financial economics does not appear to have a good explanation yet. An obvious but unsettling reason might be that they simply enjoy it. Another explanation is overconfidence, as in Odean (1998). Glode, Green, and Lowery (2010) and Bolton, Santos, and Scheinkman (2011) offer rational models where some type of informed trading might be excessive. Pagnotta and Philippon (2011) present a model where trading speed can be excessive. In their model, advances in IT do not necessarily improve the efficiency of financial markets.

7 Concluding Remarks

I have provided benchmark measures of production and efficiency for financial intermediation in the U.S. over the past 130 years. The cost of intermediation represents an annual spread of 1.5 to 2% but it has recently been trending up. This represents a puzzle. This paper does not take a stand on the likely cause of this increasing cost. Financial prices may have become more informative and risk management may have improved in such a way as to justify the increasing cost of intermediation. Alternatively, the increasing cost might reflect inefficiencies driven by zero-sum trading activities or by inefficient regulations. More research is needed to answer these important questions.

Appendices

A Notes for neoclassical Growth Model

Households

There is a representative individual who makes all the inter-temporal decisions. She owns the capital stock K_t and maximizes her expected lifetime utility

$$\mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t u(C_t) \right],$$

subject to the budget constraint $K_{t+1} + C_t \leq (1 + r_t) K_t + W_t \bar{n}_t$. From this program we obtain the well-known Euler equation $u_C(t) = \beta \mathbb{E}_t [(1 + r_{t+1}) u_C(t+1)]$. We assume constant returns to scale in production, CRRA preferences $u(C) = \frac{C^{1-\rho}-1}{1-\rho}$, and inelastic labor supply $\bar{n} = 1$. The Euler equation written with scaled quantities is then

$$\beta \mathbb{E}_t \left[(1 + r_{t+1}) (1 + \gamma_{t+1})^{-\rho} \left(\frac{c_{t+1}}{c_t} \right)^{-\rho} \right] = 1. \tag{10}$$

Production

Firms maximize profits period by period: $\{\bar{n}_t, K_t\} = \arg\max_{n,k} F(K, A_t n) - (r_t + \delta) K - W_t n$. The Cobb-Douglass production function is the special case $Y = F(K, An) = K^{1-\alpha} (An)^{\alpha}$. The two first order conditions are $F_K(t) = r_t + \delta$, and $F_N(t) = \frac{W_t}{A_t}$. Using $n = 1$, we obtain the capital demand equation

$$\frac{\partial F}{\partial k}(1, k_t) = r_t + \delta \tag{11}$$

The marginal product of capital must be equal to the rental rate plus the depreciation rate. Finally, detrended output is simply $y_t = k_t^{1-\alpha}$.

Equilibrium

The resource constraint leads to the capital accumulation equation $K_{t+1} = Y_t - C_t + (1 - \delta) K_t$, that can be written with the scaled variables:

$$(1 + \gamma_{t+1}) k_{t+1} = y_t + (1 - \delta) k_t - c_t. \tag{12}$$

The equilibrium involves the state variables $\{A_t, k_t\}$, the three unknowns endogenous variables $\{k_t, c_t, r_t\}$, and the three equilibrium conditions (10), (11), (12).

Inadequacy of Standard Multi-Sector Model

The neoclassical growth model can easily be extended to accommodate two industrial sectors. I argue, however, that the traditional multi-sector model is not useful to analyze financial intermediation. With two industrial sectors, the typical approach is to write final output as a CES aggregate of the output of the two sectors: $Y_t = \left(\omega (y_{1t})^{\frac{\sigma-1}{\sigma}} + (1-\omega)(y_{2t})^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$. The prediction of such a model are well known. The relative output of sector 1 is given by $\frac{y_1}{y_2} = \left(\frac{\omega}{1-\omega} \left(\frac{A_1}{A_2} \right)^{\alpha} \right)^{\sigma}$ in real terms, and by $\frac{p_1 y_1}{p_2 y_2} = \left(\frac{\omega}{1-\omega} \right)^{\sigma} \left(\frac{A_1}{A_2} \right)^{\alpha(\sigma-1)}$ in nominal terms. So the nominal GDP share of sector i increases with relative technological progress in sector i if and only if the elasticity of substitution is less than one. With goods and services that enter directly into the utility function, we can think of σ as a structural parameter. But this approach cannot shed light on financial intermediation because it is not the reduced form of any sensible model of financial intermediation. This will become evident in the next two sections. Section 2 introduces the simplest model of intermediation services to firms. In this model, σ is not a structural parameter. It depends both on the shape of the distribution of borrowers and on the efficiency of the supply of financial services. In other words, even in the simplest corporate finance model, σ would depend on A_1/A_2. Section 3 introduces financial services to households, and shows again that the standard multi-sector model is not useful to understand the finance industry. To summarize, it is wrong to think of some stable, demand-determined value for σ.

B Proof of Proposition 1

Let us use a production function with a constant elasticity of substitution to understand the behavior of the GDP share of finance. Output is

$$y = F(\bar{n}, k) = \left(\alpha \bar{n}^{\frac{\sigma-1}{\sigma}} + (1-\alpha) k^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}.$$

The optimality condition for factor demands are simply

$$(1-\alpha)\left(\frac{y}{k}\right)^{\frac{1}{\sigma}} = r + \delta + \psi,$$

and $\alpha\left(\frac{y}{\bar{n}}\right)^{\frac{1}{\sigma}} = w$. Therefore $\frac{k}{f(k)} = \left(\frac{1-\alpha}{r+\delta+\psi}\right)^{\sigma}$ and the finance share of GDP is

$$\phi = \left(\frac{1-\alpha}{r+\delta+\psi}\right)^{\sigma} \psi.$$

Hence $\log\phi = \sigma \log(1-\alpha) + \log\psi - \sigma \log(r+\delta+\psi)$, and

$$\frac{d\log\phi}{d\psi} = \frac{1}{\psi} - \frac{\sigma}{r+\delta+\psi}.$$

Therefore $\frac{d\log\phi}{d\psi} > 0$ if and only if $r + \delta > (\sigma - 1)\psi$. QED.

C Monitoring and Corporate Finance

The timing of actions within each period is depicted in the following table

Table 1: Events within Time t

	(i) Investment	(ii) Effort	(iii) Labor Market	(iv) Payoffs
Household (mass 1)	Capital K_t	.	Earn W_t	Earn $(1+r)K$
Entrepreneur (mass k)	Project size A_t	Effort+Monitoring	Hire n	γA
Key Equation	$K_t = kA_t$	(IC) constraint	$A_t \pi(w)$	Value: $(\pi + 1 - \delta) A_t$

There are two types of agents. Households live forever, as in the neoclassical benchmark. Entrepreneurs live for one period and have no initial wealth. The number of active entrepreneurs k is endogenous. Each entrepreneur operates a project of size A that yields a payoff of πA and a salvage value of capital of $(1-\delta)A$. Creditors must be repaid $(1+r)A$. The net present value of a project is therefore

$$(\pi + 1 - \delta) A_t - (1+r) A_t = (\pi - r - \delta) A_t$$

As before, net income is defined by $\pi(w) \equiv \max f(n) - wn$ with labor demand $n(w)$ as the solution to $f'(n) = w$.

C.1 Credit Constraints

A project is successful only if the entrepreneur behaves well. If not, the project fails, investors recover nothing, while the entrepreneur enjoys the private benefit γA. One interpretation is that the entrepreneur consumes a fraction η

of the capital stock, while the remaining $1-\eta$ is wasted. Investors must make sure that the entrepreneur behaves well. Let sA be the payoff to the entrepreneur in case of success. The incentive compatibility constraint for the entrepreneur is therefore $sA > \eta A$, or simply $s > \eta$. The pledgeable income of the project is the maximum value that the lenders can receive $(\pi - \eta) A + (1-\delta) A$. The project can be financed if and only if $\pi - \eta > r + \delta$. In the first best economy under constant returns to scale we have of course $\pi = r + \delta$. Therefore there will always be rationing relative to first best. Therefore the wage is pinned down by

$$\pi (w) = r + \delta + \eta.$$

There are several ways to model credit constraints. The important point is that punishment is limited on the downside. This implies that incentives must be provided on the upside, thereby limiting pledgeable income, and inducing credit rationing. Here punishment is bounded by assumption of limited liability. The standard moral hazard literature (e.g., Holmström (1979)) does not impose limited liability. Instead punishments are endogenously limited by the interaction between risk aversion and imperfect ex-post performance measures (captured by likelihood ratios). Of course, one can combine limited liability with imperfect performance measurement as in Holmström and Tirole (1997). Because these details are not important for my main point, I choose to work with the simpler limited liability model.

C.2 Monitoring

We can now extend the basic model by introducing intermediaries with a monitoring technology. Each entrepreneur is characterized by its monitoring need μ. Monitoring prevents the entrepreneur from enjoying private benefits. Let ψ be the unit cost of of monitoring services. The financing constraint with monitoring becomes

$$\mu\psi < \pi - r - \delta$$

Note the simplifying assumption here that μ does not depend on π (see below for model where it does). Consider now the simplest production function for financial services. Monitoring is provided by bankers, each banker can produce ψ^{-1} units of monitoring services. If an entrepreneur can get funding without monitoring it will do so. But if the funding constraint binds, monitoring is valuable in equilibrium as long as

$$\mu\psi < \gamma$$

The distribution of entrepreneurs is $G(\mu) = G_0 + \int_0^\mu g(x)\,dx$, with G_0 entrepreneurs without monitoring needs. Let us assume for simplicity that the equilibrium wage is such that $\pi(w) < r + \delta + \gamma$. This means that entrepreneurs must purchase their μ units of monitoring to obtain financing (of course, this μ is actually zero for a mass G_0 of them). The marginal type is then defined by $\pi(w) = r + \delta + \hat{\mu}\psi$ and the number of firms is $k(\hat{\mu}) = G(\hat{\mu})$. The system in $(\hat{\mu}, w)$ is then simply

$$\begin{aligned} \pi (w) &= r + \delta + \psi\hat{\mu} \\ 1 &= G(\hat{\mu})\, n(w) \end{aligned}$$

This is the system (6) and (7) in the main text. It is clear that finance is constant on the Balanced Growth Path.[23]

D Proof of Proposition 3

Clearing the intermediation market requires $A\int_0^{\hat{\mu}} \mu g(\mu)\,d\mu = F_\phi(A\bar{n}_\phi, K_\phi)$. Under HIH, we can write $F_\phi(A\bar{n}_\phi, K_\phi) = Ak_\phi f_\phi(n_\phi)$ where $n_\phi = \bar{n}_\phi / k_\phi$ is the labor-capital ratio in banking (equivalent to the number of employees per bank if we think of k_ϕ as the number of banks). Hence we can write market clearing as

$$\int_0^{\hat{\mu}} \mu g(\mu)\,d\mu = k_\phi f_\phi(n_\phi)\,. \tag{13}$$

[23]It is also straightforward to introduce moral hazard on the side of banks. In this case, banks must have incentives to monitor and they must receive a minimum amount y^μ in the good state. This is the model of Holmström and Tirole (1997).

Each banks maximize profits, therefore $n_\phi = \arg\max_n \psi f_\phi(n) - wn$, and there is free entry in banking, so $\psi f_\phi(n_\phi) - n_\phi = r + \delta$. Hence we get

$$\psi f'_\phi(n_\phi) = w \tag{14}$$

and $\psi f_\phi(n_\phi) = wn_\phi + r + \delta$, which we can combine into

$$\frac{f'_\phi(n_\phi)}{f_\phi(n_\phi)} = \frac{w}{wn_\phi + r + \delta} \tag{15}$$

This last equation (15) pins downs the function $n_\phi(w)$. Then we get the marginal cost $\psi(w)$ from (14). Of course in the real sector we have

$$\pi(w) = r + \delta + \psi(w)\hat{\mu} \tag{16}$$

and we clear the labor market with

$$1 = G(\hat{\mu})\, n(w) + k_\phi n_\phi(w) \tag{17}$$

The three unknowns are $(w, k_\phi, \hat{\mu})$ and the three equilibrium conditions are (13), (16) and (17). QED.

E Household Credit Accounting

Capital market clearing requires

$$S_t = K_{t+1} + B_t$$

Adding up the budget constraints we have

$$W_t + (1 + r_t)\, S_{t-1} + (1 - \psi_c)\, B_t - (1 + r_t)\, B_{t-1} = C_{st} + C_{1t} + C_{2t} + S_t$$

The two sides of GDP are

$$\begin{aligned} Y_t &= W_t + (r_t + \delta + \psi_k)\, K_t \\ Y_t &= K_{t+1} + C_{st} + C_{1t} + C_{2t} - (1 - \delta - \psi_k)\, K_t + \psi_c B_t \end{aligned}$$

Combining them we get

$$K_{t+1} + C_{st} + C_{1t} + C_{2t} = W_t + (1 + r_t)\, K_t - \psi_c B_t$$

Combining with the budget constraints we get

$$(1 - \psi_c)\, B_t - B_t = -\psi_c B_t$$

which is simply the zero profit condition for consumer credit intermediaries.

F Corporate Quality Adjustment

Here I present a special model where heterogeneity in monitoring requirements stems from differences in free cash. Let us thus assume that the entrepreneur has cash-on-hand X. The pledgeable income is still $(\pi - \eta)\, A + (1 - \delta)\, A$ so the funding constraint without monitoring becomes $(\pi - \eta)\, A + (1 - \delta)\, A > (1 + r)\,(A - X)$, i.e., entrepreneurs are credit-constrained as long as

$$(1 + r)\, x < r + \delta - \pi + \eta$$

In this case, the entrepreneur needs to hire monitoring services. As before, suppose that hiring μ costs $\psi\mu$ and reduces private benefits from η to $\eta - \mu$. Suppose that there are two types of borrowers, $i = l, h$ with $\Delta \equiv x_h - x_l > 0$. In equilibrium, they will satisfy their funding requirements with equality by demanding

$$\mu_i = r + \delta - \pi + \eta - (1 + r)\, x_i$$

Notice the key intuition. The first few dollars of borrowing do not require any monitoring. So monitoring requirements are a convex function of borrowing (leverage).

Suppose that there are k borrowers and a fraction λ are low cash. Aggregate monitoring is

$$\bar{\mu} = k\left(\left(1-\lambda\right)\mu_h + \lambda\mu_l\right) = k\left(r+\delta-\pi+\eta-\left(1+r\right)x_h\right) + \left(1+r\right)\lambda k\Delta$$

Aggregate borrowing is

$$\bar{b} = k\left(1-x_h+\lambda\Delta\right)$$

Monitoring intensity, defined as the ratio of monitoring to borrowing, is

$$\frac{\bar{\mu}}{\bar{b}} = \frac{r+\delta-\pi+\eta-\left(1+r\right)x_h+\left(1+r\right)\lambda\Delta}{1-x_h+\lambda\Delta}$$

This formula provides a way to make a quality adjustment to the corporate output series as a function of λ.

G Complete Structural Model

This is the model with household credit and heterogenous corporate borrowers.

G.1 Firms

With Cobb-Douglass technology, we get net income

$$\begin{aligned}
\pi\left(w\right) &= \left(1-\alpha\right)\left(\frac{\alpha}{w}\right)^{\frac{\alpha}{1-\alpha}} \\
n &= \left(\frac{\alpha}{w}\right)^{\frac{1}{1-\alpha}}
\end{aligned}$$

Monitoring demand

$$\begin{aligned}
\mu_l &= r+\delta-\pi\left(w\right)+\eta-\left(1+r\right)x_l \\
\mu_h &= r+\delta-\pi\left(w\right)+\eta-\left(1+r\right)x_h
\end{aligned}$$

Marginal firm low type with free entry: $\pi\left(w\right) = r+\delta+\psi_k\mu_l$, which we can write $\pi\left(w\right) = r+\delta+\psi_k\left(r+\delta-\pi\left(w\right)+\eta-\left(1+r\right)x_l\right)$ and finally

$$\pi\left(w\right) = \bar{\pi} \equiv r+\delta+\frac{\psi_k}{1+\psi_k}\left(\eta-\left(1+r\right)x_l\right)$$

This pins down w, then w pins down n, then we get the number of firms from the labor market clearing condition

$$kn\left(w\right) = 1$$

Let k_h be the (exogenous) number of high cash firms (these are infra-marginal, equivalent to G_0 in Section 2). Then we get

$$k_l = k - k_h$$

We need to check that $k_l > 0$ and that $\mu_h > 0$. And we get aggregate corporate monitoring as

$$\begin{aligned}
\bar{\mu}_k &= k_l\mu_l + k_h\mu_h = \mu_l k - k_h\left(\mu_l-\mu_h\right) \\
\mu_l-\mu_h &= \left(1+r\right)\left(x_h-x_l\right)
\end{aligned}$$

and corporate borrowing

$$\bar{b}_k = k - k_l x_l - k_h x_h$$

Output is simply

$$y = k^{1-\alpha}$$

And corporate finance income is

$$\phi_k = \psi_k \frac{\bar{\mu}_k}{y}$$

Using Cobb-Douglass, we have

$$\bar{\pi} \equiv r + \delta + \frac{\psi_k}{1+\psi_k} \left(\eta - (1+r)\, x_l\right)$$

$$w = \alpha \left(\frac{1-\alpha}{\bar{\pi}}\right)^{\frac{1-\alpha}{\alpha}} = \alpha y$$

$$n = \left(\frac{\bar{\pi}}{1-\alpha}\right)^{\frac{1}{\alpha}}$$

$$k = \left(\frac{1-\alpha}{\bar{\pi}}\right)^{\frac{1}{\alpha}}$$

$$\frac{k}{y} = k^{\alpha} = \frac{1-\alpha}{\bar{\pi}}$$

$$\mu_l = \frac{\eta - (1+r)\, x_l}{1+\psi_k}$$

$$\bar{\mu}_k = \mu_l k - (1+r)\left(x_h - x_l\right) k_h$$

G.2 Households

Long-term households pin down the interest rate

$$\begin{aligned} \theta &\equiv \rho - \nu\left(1-\rho\right) \\ \beta\left(1+r\right) &= \left(1+\gamma\right)^{\theta} \end{aligned}$$

Money demand is

$$\psi_m m = \nu c$$

Short-lived households consumption equation is such that

$$c_1 = \left(1-\psi_c\right)^{\frac{1}{\theta}} c_2.$$

$(1+\nu)\, c_1 = \eta_1 w + (1-\psi_c)\, b$ and $(1+\nu)\, c_2 = \eta_2 w - \frac{1+r}{1+\gamma} b$.

Using $w = \alpha y$:

$$\frac{b}{y} = \frac{\left(1-\psi_c\right)^{\frac{1}{\theta}} \eta_2 - \eta_1}{1 - \psi_c + \left(1-\psi_c\right)^{\frac{1}{\theta}} \frac{1+r}{1+\gamma}} \alpha.$$

To close the model, we obtain the consumption of the long lived savers (c_0) from the resource constraint:

$$y = (1+\nu)\left(c_0 + c_1 + c_2\right) + \psi_c b + (\gamma + \delta)\, k + \psi_k \bar{\mu}_k.$$

Since $\eta_1 + \eta_2 = 1$, and $w = \alpha y$, we know that $(1+\nu)\left(c_1 + c_2\right) = \alpha y + (1-\psi_c)\, b - \frac{1+r}{1+\gamma} b = \alpha y - \left(\psi_c + \frac{r-\gamma}{1+\gamma}\right) b$

therefore

$$(1+\nu)\,c_0 = (1-\alpha)\,y + \frac{r-\gamma}{1+\gamma}b - (\gamma+\delta)\,k - \psi_k \bar{\mu}_k$$

Since $(1-\alpha)\,y = \bar{\pi}k$ we have

$$(1+\nu)\,c_0 = (r-\gamma)\left(k + \frac{b}{1+\gamma}\right) + \frac{\psi_k}{1+\psi_k}\left(\eta - (1+r)\,x_l\right)k - \psi_k \bar{\mu}_k$$

From monitoring and free entry we know that $(1+\psi_k)\,\mu_l = \eta - (1+r)\,x_l$ and $\bar{\mu}_k = k_l \mu_l + k_h \mu_h = \mu_l k - k_h\,(\mu_l - \mu_h)$, therefore

$$\frac{\psi_k}{1+\psi_k}\left(\eta - (1+r)\,x_l\right)k - \psi_k \bar{\mu}_k = \psi_k k_h\,(\mu_l - \mu_h) = (1+r)\,\psi_k k_h\,(x_h - x_l)$$

and therefore we get

$$(1+\nu)\,c_0 = (r-\gamma)\left(k + \frac{b}{1+\gamma}\right) + (1+r)\,\psi_k k_h\,(x_h - x_l)$$

Total expenditure of long-lived households is equal to their capital income from loans to corporates and to short-lived households, as well as the pure profits earned by firms with low monitoring requirements. The GDP share of finance is

$$\phi = \psi_m \frac{m}{y} + \psi_c \frac{b_c}{y} + \psi_k \frac{\bar{\mu}_k}{y}.$$

I have not discussed in great details how managers and long term households are related. Incentives constraints refer to managerial decisions. The simplest way to rationalize the accounting of income in the model is to follow Gertler and Karadi (2009) and Gertler and Kiyotaki (2010).

Household adjustment [to be completed]

There are many ways to model heterogeneity. Here I use a simple model where the ex-post income of households is random. Half of the households draw a relatively high income η_2^h and the other half a low income η_2^l. To implement the allocation there must be transfers across households equal to

$$\tilde{b}_2 = \frac{\eta_2^h - \eta_2^l}{2}$$

Assuming these cost the same as the initial loans we get

$$\psi_c\left(\frac{b}{y} + \frac{\eta_2^h - \eta_2^l}{2y}\right)$$

$$C_{2t+1} + \psi_m M_{2t+1} = \eta_2 W_{t+1} - (1+r_{t+1})\,B_t$$

H JP Morgan 2010

According to its 2010 annual, total net revenue for JPM Co was \$103 billion, 51b of interest income and 52b of non-interest income. The investment bank earned \$26 billion, 15 from fixed income markets, 5 from equity markets, and a bit more than 6 in fees. Of the 26, non interest income accounted for 18, including 6.2b in fees (3.1 and 1.6 for debt and equity underwriting, and 1.5 for advisory fees), 8.4b from principal transactions, and 2.5b from asset management fees. For its private clients, the investment bank raised 440b in debt and 65b in equity. This suggests underwriting fees of $3.1/440 = 0.70\%$ for debt, and $1.6/65 = 2.46\%$ for equity. The cost of equity underwriting is therefore about 3.5 times the cost of debt underwriting. The bank also raised 90b for US govt and non-profits. The bank advised 311 announced M&A (a 16% market share). The bank also loaned or arranged 350b.

References

Acharya, V. V., L. H. Pedersen, T. Philippon, and M. Richardson (2009): "Measuring Systemic Risk," Working Paper NYU.

Adrian, T., and H. S. Shin (2008): "Financial Intermediary Leverage and Value at Risk," Federal Reserve Bank of New York Staff Reports, 338.

Aguiar, M., and M. Bils (2011): "Has Consumption Inequality Mirrored Income Inequality?," Working Paper, University of Rochester.

Altinkilic, O., and R. S. Hansen (2000): "Are There Economies of Scale in Underwriting Fees? Evidence of Rising External Financing Costs," *Review of Financial Studies*, 13, 191–218.

Bai, J., T. Philippon, and A. Savov (2011): "Have Financial Prices Become More Informative?," .

Baker, M., and J. Wurgler (2000): "The Equity Share in New Issues and Aggregate Stock Returns," *Journal of Finance*.

Basu, S., R. Inklaar, and J. C. Wang (2011): "The Value of Risk: Measuring the Services of U.S. Commercial Banks," *Economic Inquiry*, 49(1), 226–245.

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.

Bickenbach, F., E. Bode, D. Dohse, A. Hanley, and R. Schweickert (2009): "Adjustment After the Crisis: Will the Financial Sector Shrink?," Kiel Policy Brief.

Bolton, P., T. Santos, and J. Scheinkman (2011): "Cream Skimming in Financial Markets," Working Paper, Columbia University.

Carter, S., S. Gartner, M. Haines, A. Olmstead, R. Sutch, and G. Wright (eds.) (2006): *Historical Statistics of the United States Millennial Edition Online*Cambridge University Press.

Chirinko, R. S., S. M. Fazzari, and A. P. Meyer (2011): "A New Approach to Estimating Production Function Parameters: The Elusive Capital–Labor Substitution Elasticity," *Journal of Business and Economic Statistics*, 29(4), 587–594.

Diamond, D., and R. Rajan (2001): "Liquidity Risk, Liquidity Creation and Financial Fragility: A Theory of Banking," *Journal of Political Economy*, 109, 287–327.

Estevadeordal, A., B. Frantz, and A. M. Taylor (2003): "The Rise and Fall of World Trade, 1870-1939," *The Quarterly Journal of Economics*, 118(2), 359–407.

French, K. R. (2008): "Presidential Address: The Cost of Active Investing," *The Journal of Finance*, 63(4), 1537–1573.

Gennaioli, N., A. Shleifer, and R. Vishny (2011): "A Model of Shadow Banking," Working Paper, Harvard University.

Gerardi, K. S., H. Rosen, and P. Willen (2010): "The Impact of Deregulation and Financial Innovation on Consumers: The Case of the Mortgage Market," *Journal of Finance*, 65(1), 333–360.

Gertler, M., and P. Karadi (2009): "A Model of Unconventional Monetary Policy," Working Paper NYU.

Gertler, M., and N. Kiyotaki (2010): "Financial Intermediation and Credit Policy in Business Cycle Analysis," Working Paper, NYU.

Glode, V., R. C. Green, and R. Lowery (2010): "Financial Expertise as an Arms Race," forthcoming in Journal of Finance.

Gorton, G., S. Lewellen, and A. Metrick (2012): "The Safe-Asset Share," NBER Working Paper.

Gorton, G., and G. Pennacchi (1990): "Financial Intermediaries and Liquidity Creation," *Journal of Finance*, 45(1), 49–72.

Gorton, G., and A. Winton (2003): "Financial Intermediation," in *Handbook of the Economics of Finance*, ed. by G. M. Constantinides, M. Harris, and R. Stulz, pp. 431–552, North Holland. Elsevier.

Gorton, G. B., and A. Metrick (2012): "Securitized Banking and the Run on Repo," *Journal of Financial Economics*.

Greenwood, R., S. Hanson, and J. Stein (2011): "A Comparative-Advantage Approach to Government Debt Maturity," Working Paper Harvard University.

Guiso, L., P. Sapienza, and L. Zingales (2004): "The Role of Social Capital in Financial Development," *American Economic Review*, 94, 526–556.

Hadas, E. (2011): "Commodity prices are failing New Zealand test," *Financial Times*.

Haldane, A., S. Brennan, and V. Madouros (2010): "What is the contribution of the financial sector: Miracle or mirage?," in *The Future of Finance: The LSE Report*, ed. by A. T. et al. LSE.

Hasbrouck, J. (2009): "Trading Costs and Returns for U.S. Equities: Estimating Effective Costs from Daily Data," *Journal of Finance*, 64(3), 1445–1477.

Holmström, B. (1979): "Moral Hazard and Observability," *Bell Journal of Economics*, 10(1), 74–91.

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

Jovanovic, B., and P. L. Rousseau (2001): "Why Wait? A Century of Life Before IPO," *AER paper and proceedings*, 91, 336–341.

Jovanovic, B., and P. L. Rousseau (2005): "General Purpose Technologies," in *Handbook of Economic Growth*, ed. by P. Aghion, and S. Durlauf. Elsevier, Chapter 18.

Kashyap, A., R. Rajan, and J. Stein (2002): "Banks as Liquidity Providers: An Explanation for the Coexistence of Lending and Deposit-Taking," *Journal of Finance*, 57, 33–73.

Krishnamurthy, A. (2009): "Amplification Mechanisms in Liquidity Crises," *NBER Working Paper*.

Krishnamurthy, A., and A. Vissing-Jorgensen (2010): "The Aggregate Demand for Treasury Debt," Working Paper Kellogg School of Management.

Kuznets, S. (1941): "National Income and Its Composition, 1919-1938," Discussion paper, National Bureau of Economic Research.

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

Lucas, R. E. (1980): "Equilibrium in a Pure Currency Economy," in *Models of Monetary Economics*, ed. by J. H. Karaken, and N. Wallace. Federal Reserve Bank of Minneapolis.

Lucas, R. E. J. (2000): "Inflation and Welfare," *Econometrica*, 68(2), 247–274.

Lucas, R. E. J., and N. L. Stokey (1987): "Money and Interest in a Cash-in-Advance Economy," *Econometrica,*, 55(3), 491–513.

Martin, R. F. (1939): *National Income in the United States, 1799-1938*. National Industrial Conference Board,.

Mayer, C., and K. Pence (2008): "Subprime Mortgages: What, Where, and to Whom?," Staff Paper Federal Reserve Board.

Mehra, R., F. Piguillem, and E. C. Prescott (2011): "Costly financial intermediation in neoclassical growth theory," *Quantitative Economics*, 2(1), 1–36.

Mehra, R., and E. C. Prescott (1985): "The Equity Premium: a Puzzle," *Journal of Monetary Economics*, 15, 145–161.

Merton, R. C. (1995): "A Functional Perspective of Financial Intermediation," *Financial Management*, 24, 23–41.

Moore, K. B., and M. G. Palumbo (2010): "The Finances of American Households in the Past Three Recessions: Evidence from the Survey of Consumer Finances," Staff Paper Federal Reserve Board.

Obstfeld, M., and A. M. Taylor (2002): "Globalization and Capital Markets," NBER Working Paper 8846.

Odean, T. (1998): "Volume, Volatility, Price, and Profit when all traders are above average," *Journal of Finance*, 53, 1887–1934.

O'Mahony, M., and M. P. Timmer (2009): "Output, Input and Productivity Measures at the Industry Level: The EU KLEMS Database," *The Economic Journal*, 119(538), F374–F403.

Pagnotta, E., and T. Philippon (2011): "Competing on Speed," Working Paper NYU.

Philippon, T. (2008): "The Evolution of the US Financial Industry from 1860 to 2007: Theory and Evidence," NBER Working Paper No. 13405.

——— (2012): "Finance vs. Wal-Mart: Why are Financial Services so Expensive?," in *Lessons from the Crisis*, ed. by A. Blinder, A. Lo, and R. Solow.

Philippon, T., and A. Reshef (2007): "Wages and Human Capital in the U.S. Financial Industry: 1909-2006," NBER Working Paper 13437.

Piketty, T., and E. Saez (2003): "Income Inequality in the United States, 1913-1998," *Quarterly Journal of Economics*, 118(1), 1–39.

Pozsar, Z., T. Adrian, A. Ashcraft, and H. Boesky (2010): "Shadow Banking," NY Fed Staff Report.

Rajan, R. G., and L. Zingales (1998): "Financial Dependence and Growth",," *American Economic Review*, 88, 559–586.

Raurich, X., H. Sala, and V. Sorolla (2011): "Factor Shares, the Price Markup, and the Elasticity of Substitution between Capital and Labor," IZA Working Paper.

Reinhart, C. M., and K. S. Rogoff (2009): *This Time Is Different: Eight Centuries of Financial Folly.* Princeton University Press.

Rhodes-Kropf, M., and D. T. Robinson (2008): "The Market for Mergers and the Boundaries of the Firm," *The Journal of Finance*, 63(3), 1169–1211.

Ritter, J. R. (2011): "Initial Public Offerings," mimeo available online.

Sargent, T. J., and B. D. Smith (2009): "The Timing of Tax Collections and the Structure of "Irrelevance" Theorems in a Cash-in-Advance Model," mimeo NYU.

Stein, J. (2012): "Monetary Policy as Financial-Stability Regulation," *The Quarterly Journal of Economics*, 127(1), 57–95.

Tang, K., and W. Xiong (2011): "Index Investment and Financialization of Commodities," .

Wang, J. C., S. Basu, and J. G. Fernald (2008): "A General-Equilibrium Asset-Pricing Approach to the Measurement of Nominal and Real Bank Output," NBER Working Paper No. 14616.