Has the U.S. Finance Industry Become Less Efficient?

On the Theory and Measurement of Financial Intermediation

Thomas Philippon*

October 2013

Abstract

A quantitative investigation of financial intermediation in the U.S. over the past 130 years yields the following results : (i) the finance industry's share of GDP is high in the 1920s, low in the 1950s and 1960s, and high again in the 1990s and 2000s; (ii) most of these variations can be explained by corresponding changes in the quantity of intermediated assets (equity, household and corporate debt, assets yielding liquidity services); (iii) intermediation is produced under constant returns to scale with an annual average cost comprised between 1.5% and 2% of outstanding assets; (iv) quality adjustments that take into account changes in the characteristics of firms and households are quantitatively important; and (v) the unit cost of intermediation has not decreased over the past 30 years.

JEL: E2, G2, N2

*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. Darrell Duffie, Robert Lucas, Raghuram Rajan, Jose Scheinkman, Robert Shiller, Andrei Shleifer, and Richard Sylla have provided invaluable feedback at various stages of this project. Boyan Jovanovic, Peter Rousseau, Moritz Schularick, and Alan Taylor have shared their data and their insights, and I have greatly benefited from discussions with Lewis Alexander, Patrick Bolton, Markus Brunnermeier, John Cochrane, Kent Daniel, Douglas Diamond, John Geanakoplos, Gary Gorton, Robin Greenwood, Steve Kaplan, Anil Kashyap, Ashley Lester, Andrew Lo, Guido Lorenzoni, Andrew Metrick, William Nordhaus, Matthew Rhodes-Kropf, David Robinson, Kenneth Rogoff, Tano Santos, David Scharfstein, Hyun Shin, Jeremy Stein, Gillian Tett, Wallace Turbeville, and Luigi Zingales, as well as seminar participants at Stanford, Yale, NYU, Harvard, Chicago, Princeton, and the Paris School of Economics. I also thank Paul Krugman for his discussion at the 2011 NY Area Monetary conference, Axelle Ferrière, Peter Gross, Andrea Prestipino, Robert Turley, 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. 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 aggregate 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. This cost of financial intermediation is reflected in the user cost of external finance for firms who need to issue debt and equity, and for households who want to borrow or use asset management services. In equilibrium, the user cost of external finance is the sum of the rate of returns to saver (r) and the unit cost of financial intermediation (ψ):

$$\text{user cost of finance} = r + \psi. \tag{1}$$

The unit cost of intermediation ψ can in turn be measured as the ratio of the income of financial intermediaries to the quantity of intermediated assets. The goal of this paper is to construct these three measures: the income of financial intermediaries, the quantity of intermediated assets, and, finally, the unit cost of intermediation ψ.

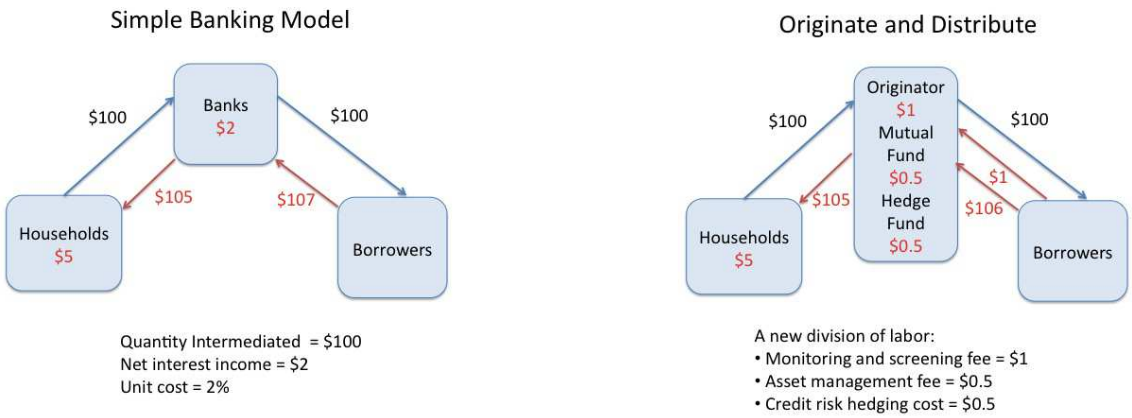
There are several motivations for undertaking such a project. A first motivation is that ψ has a direct impact on the overall efficiency of the economy. Even small changes in ψ have large long run effects on the capital/output ratio, and therefore on income per capita. Equations such as (1) play a central role in the literature that seeks to quantify the consequences of financial development for economic growth.[1]

A second motivation is to shed light on the transformation of the finance industry that has occurred since the 1970s. For instance, we would certainly like to know if the move away from traditional banking and towards an "originate-and-distribute" model has lowered the cost of funds for households and businesses. This is precisely what ψ should measure. The debate about capital adequacy ratios for banks is also a debate about ψ. Bank leverage has risen substantially since the late 19th century, as discussed in Haldane et al. (2010) among others. If Modigliani and Miller (1958)'s proposition holds, as Admati et al. (2011) argue, we should not expect a link between ψ and bank leverage, but if the proposition fails we might expect a downward drift in ψ over time. Similarly, if derivatives markets lower hedging costs, their growth should translate into lower funding costs and higher assets values. The broader point here is that learning about ψ is important from a positive perspective and from a normative perspective.

This paper seeks to define and measure financial intermediation. It treats the finance industry as a black box and attempts to measure what goes in, what comes out, and how much the whole system costs. It is important,

[1] See Greenwood et al. (2010), Buera et al. (2011), and Midrigan and Xu (2011) for recent analyses of financial development and growth. In addition, much recent work has focused the macroeconomic consequences of a sudden increase in ψ, and on the link between ψ and intermediary capital, leverage and liquidity. Curdia and Woodford (2009), Gertler and Kiyotaki (2010), Hall (2011), Christiano and Ikeda (2011), Corsetti et al. (2011) study the impact of negative shocks to financial intermediation, building on the classic contribution of Bernanke et al. (1999). Gertler and Karadi (2011), He and Krishnamurthy (2012), and Moore (2011) focus on liquidity. This paper only deals with the long term evolution of ψ, but the value of ψ in normal times is an important parameter even if one is interested in the deviations from its long term trend.

Figure 1: Two Equivalent Models of Financial Intermediation

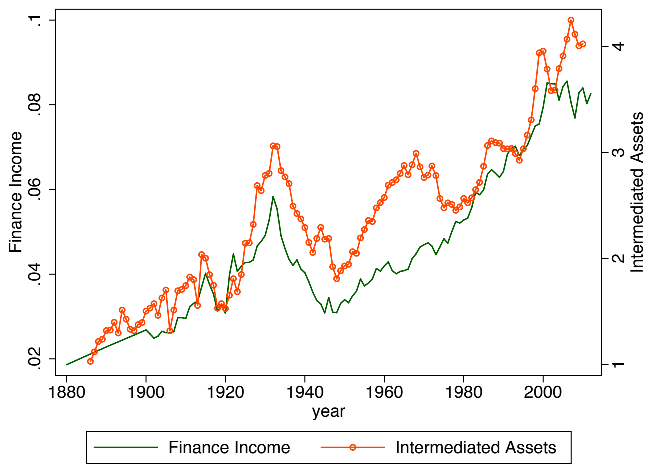


however, to understand just how difficult the measurement problem is. A simple illustration is given in Figure 1. At the prevailing market rates of 5% and 7%, borrowers (firms or households) want to borrow \$100, and savers want to save \$100. To flow back and forth between savers and borrowers, the funds go through financial intermediaries. These intermediaries need \$2 to pay their wage bill and rent the necessary capital. In the terminology of this paper, the quantity of intermediated assets is 100, the intermediation cost is 2, and therefore $\psi = 2\%$. Figure 1 presents two fundamentally equivalent ways to organize financial intermediation. In traditional banking, intermediation occurs under one roof: the bank makes a loan, keeps it on its books, and earns a net interest income. This income compensates for the cost of screening and monitoring the borrower, and for managing the duration and credit risk of the loan. In the originate and distribute model, by contrast, there is a daisy chain of intermediation. Many transactions occur *inside* the black box, with face values potentially much larger than \$100. There is no simple measure of net interest income as in the traditional model: there are origination fees, asset management fees, trading profits, etc. But the sum of wages and profits for all intermediaries is still \$2, and the quantity of assets intermediated seen from *outside* the black box, is still \$100.

Three steps are required, then, to understand financial intermediation, and these steps determine the structure of the paper: (i) measure the income of financial intermediaries; (ii) define and construct the quantity of intermediated assets; and (iii) compute the unit cost of intermediation and perform quality adjustments. For the purposes of this paper, the real difficulty lies in the heterogeneity among intermediated assets. On the liability side of intermediaries, savings vehicles are heterogenous: perhaps households save \$50 in liquid assets with an interest of 4% and \$50 in illiquid assets with an interest of 6%. On the asset side of intermediaries we find debt and equity, and, most importantly, we find different types of borrowers: young firms and blue chip companies, wealthy households and poor households. Changes in the composition of borrowers affect the cost of intermediation, while improvements in financial intermediation give access to credit to borrowers who were previously priced out.

The first contribution of this paper is empirical. Figure 2 shows that the degree of financial intermediation varies dramatically over time. The first series, constructed in Section 1, is the income of financial intermediaries divided by GDP. 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% after 2000.

Figure 2: Finance Income and Intermediated Assets



Notes: Both series are expressed as a share of GDP, excluding defense spending. Finance Income is the domestic income of the finance and insurance industries, i.e., aggregate income minus net exports. Intermediated Assets include debt and equity issued by non financial firms, household debt, and various assets providing liquidity services.

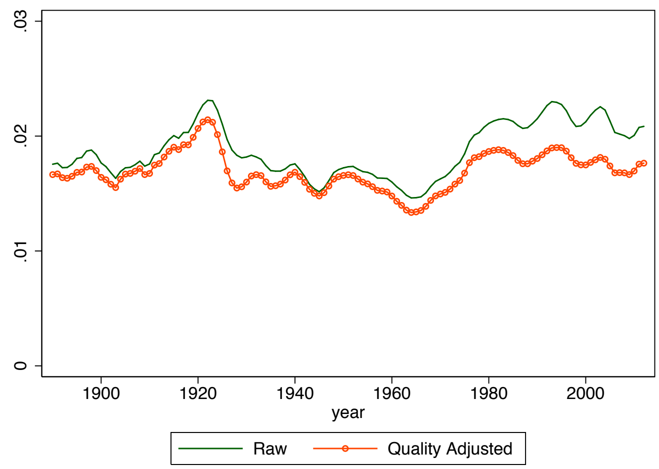
Given these large historical variations in the finance income share, it is natural to ask if there are commensurate changes in the quantity of intermediated assets.[2] Section 2 uses a simple extension of the neoclassical growth model as an accounting framework for household finance, corporate finance, and liquidity provision. The size of the various markets varies significantly over time. The most important trend in credit markets in recent years is the increase in household debt. The business credit market is relatively large in the 1920s, small in the 1960s and large again after 1980, although not as large as in the late 1920s. I also measure the market value of outstanding equity and the flows of initial and seasoned offerings. Deposits, repurchase agreements, and money markets mutual funds are used to measure liquidity services. After aggregating the various types of credit, equity issuances and liquid assets into one measure, I obtain the quantity of financial assets intermediated by the financial sector for the non-financial sector, displayed in Figure 2.

I can then divide the income of the finance industry by the quantity of intermediated assets to obtain a measure of the unit cost ψ. Figure 3 shows that this unit cost is around 2% and relatively stable over time. In other words, I estimate that it costs two cents per year to create and maintain one dollar of intermediated financial asset. I also

[2] We can think of the finance industry as providing three types of services: (i) liquidity (means of payments, cash management); (ii) transfer of funds (pooling funds from savers, screening and monitoring borrowers); (iii) information (price signals, advising on M&As). Financial firms typically produce a bundle of such services. For instance, risk management uses all three types of activities. Services of type (i) and (ii) typically involve the creation of various financial assets and liabilities. This classification is motivated by the mapping between theory and measurement discussed throughout the paper. It differs a little bit from that of Merton (1995). I do not attempt in this paper to measure the informativeness of prices. This issue is tackled by Bai et al. (2011). See the discussion at the end of Section 3.

find clear evidence that financial services are produced under constant returns to scale. For instance, from 1947 to 1973 (a period of stable growth without major financial crises), real income per-capita increases by 80% and real financial assets by 250%, but my estimate of the unit cost of intermediation remains remarkably constant.

Figure 3: Unit Cost of Financial Intermediation



Notes: The raw measure is the ratio of finance income to intermediated assets, displayed in Figure 2. The quality adjusted measure takes into account changes in firms' and households' characteristics.

The raw measure of Figure 3, however, does not take into account changes in the characteristics of borrowers. The final contribution of the paper is to perform quality adjustments to the quantity of intermediated assets. The 1920s and 1990s are times of entry by young and risky firms, and Jovanovic and Rousseau (2005) have shown that this pattern is related to waves of technological innovation. In the household credit market, relatively poor households have gained access to credit in recent years. In both cases, the challenge is to account for the fact that these borrowers require more intermediation per unit of credit extended. I rely on theory to make the required quality adjustments, which appear to be quantitatively important. According to my calculations, in the 1990s, the raw measure of intermediation underestimates the true quantity by about 25%. Given the size of intermediated markets, the failure to adjust for quality would represent a measurement error of the order of one GDP. Figure 3 shows that the adjusted unit cost is more stable than the unadjusted one.

Even with the quality adjustment, however, I find that the unit cost of intermediation has increased since the mid 1970s and is now significantly higher than it was in the 1960s and about as high as it was at the turn of the 20th century. It seems that improvements in information technologies over the past 30 years have not necessarily led to a decrease in the unit cost of intermediation. Explaining this puzzle is an active area of research, some of which is discussed at the end of Section 3.

Related literature

Financial intermediation does not have a benchmark quantitative model in the way asset pricing does. By using a model to interpret long time series of prices and quantities, and by providing a set of stylized facts for future research, this paper shares the spirit of Mehra and Prescott (1985). But because financial intermediation is a more heterogenous field than asset pricing, this paper has to draw from several strands of the literature in finance and monetary economics.

The first strand is the theory of banking and financial intermediation. While stylized and focused on macroeconomic predictions, the model developed below is consistent with leading theories of financial intermediation, such as Diamond and Dybvig (1983), Diamond (1984), Gorton and Pennacchi (1990), Holmström and Tirole (1997), Diamond and Rajan (2001), and Kashyap et al. (2002). Gorton and Winton (2003) provide a review of the literature on financial intermediation. The focus of this paper differs from that 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 do not have room to 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 itself (e.g. La Porta et al. (1998), Guiso et al. (2004)). The literature typically focuses on corporate finance (Greenwood et al. (2010), Buera et al. (2011), Midrigan and Xu (2011)), except Mehra et al. (2011) who study intermediation in a model where households save for retirement over an uncertain lifetime.[3] This paper is more closely related to a recent branch of the literature that seeks to provide risk-adjusted measures of financial productivity (Wang et al. (2008), Haldane et al. (2010), Basu et al. (2011)).

In its account of liquidity services provided by the finance industry, this paper is also related to the classic literature on money and banking. Lucas (2000) provides an analysis of money demand. Kiyotaki and Moore (2008) study the interaction of liquidity, asset prices and aggregate activity. A recent branch of this literature has focused on the rise of market-based intermediation, also called shadow banking. Pozsar et al. (2010) describe the structure of shadow banking. Gorton and Metrick (2012), Stein (2012), Gorton et al. (2012), and Gennaioli et al. (2013a) emphasize the importance of investors' demand for safe assets as a driver of shadow banking activity.

Finally, there is an emerging literature on the growth of the finance industry.[4] Philippon and Reshef (2012) share the historical perspective of this paper but focus on the composition of the finance labor force. Kaplan and

[3] My approach is complementary to this literature and uses many of its important insights. The difference is that I focus on the evolution of the entire U.S. finance industry. As a result, both theory and measurement must be expanded. For instance, following Beck et al. (2011), the literature uses cross-country data on interest-rate spreads to estimate financing frictions, e.g., Greenwood et al. (2012). To study the US finance industry, it is important to recognize that non-interest income (fees, trading revenues, etc.) is now the dominant source of income for financial firms (even for banks: see JPMorgan's 2010 annual report for instance), that consumer credit is at least as important as corporate credit, and that the shadow banking's creation of safe assets is driven by investors' liquidity demand (all these points are discussed in details below).

[4] 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. 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.

Rauh (2010) study the compensation of the highest income earners. Greenwood and Scharfstein (2013) provide an illuminating study of the growth of modern finance in the U.S. They show that two activities account for most of this growth over the past 30 years: asset management and the provision of household credit. For asset management, they uncover an important stylized fact: individual fees have typically declined but the allocation of assets has shifted towards high fee managers in such a way that the average fee per dollar of assets under management has remained roughly constant. While most of the existing work has focused on the U.S., Philippon and Reshef (2013) and Bazot (2013) provide evidence for other countries.

A second set of papers offers theoretical explanations for the growth of finance documented in this paper and in the empirical papers discussed above. There are two main stylized facts to explain: the size of finance (see figure 2), and the unit cost (see figure 3). Regarding this second stylized fact, a puzzle seems to be that the unit cost has not declined despite obvious improvements in information technologies. As a result, the income received by financial intermediaries might be unexpectedly high. In Glode et al. (2010), an "arms' race" can occur as agents try to protect themselves from opportunistic behavior by (over)-investing in financial expertise. In Bolton et al. (2011), cream skimming in one market lowers assets quality in the other market and allows financial firms to extract excessive rents. In Pagnotta and Philippon (2011) there can be excessive investment in trading speed because speed allows trading venues to differentiate and charge higher prices. Gennaioli et al. (2014) propose an alternative interpretation for the relatively high cost of financial intermediation. In their model, trusted intermediaries increase the risk tolerance of investors, allowing them to earn higher returns. Because trust is a scarce resource, improvements in information technology do not necessarily lead to a lower unit cost.

The other fact to explain is the size of the finance industry. Since the unit cost appears to be roughly constant, the question becomes: how do we explain the large historical variations in the ratio of intermediated assets over GDP? This paper documents that the income share of the finance industry is roughly equal to 2% of the ratio of intermediated assets over GDP, but it does seek to explain the size of intermediated assets.[5] Gennaioli et al. (2013b) propose an explanation. They argue that the growth of finance can be explained by the rise of the wealth to income ratio, documented in Piketty and Zucman (2013) for several countries. The driving force is a slowdown in aggregate growth which leads, along the transition path, to an increase in the capital output ratio. If the unit cost of intermediation does not fall as the capital output ratio increases, then the income share of the finance industry will increase.

Let me end this introduction with an important caveat: this paper does not analyze financial crises. The model assumes that credit markets clear via prices, and not via covenants or quantity restrictions as we often see during crises. In the model, borrowers can be priced out, but inefficient rationing does not occur. Similarly, the model does not study whether borrowing is appropriate or excessive, whether financial intermediaries take on too much

[5] The household credit model of Section 3 can "account" for some (but not all) of the rise in consumer debt due to improvements in access to credit. But even there, the goal is not to explain the size of the market, but rather to refine the measurement of the unit cost by removing the bias created by time varying fixed costs.

aggregate risk, and whether government interventions create moral hazard.[6]

The remainder of the paper is organized as follows. Section 1 estimates the income of financial intermediaries. Section 2 computes the quantity of intermediated assets. Section 3 implements quality adjustments and discusses the role of information technology and price informativeness. Section 4 concludes.

1 Income Share of Finance

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 simple way to break down the income earned by the finance industry into economically meaningful components. For instance, 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. From a historical perspective, these issues are compounded by regulatory changes in the range of activities that certain intermediaries can provide. Rather than imposing arbitrary interpretations on the data, I therefore focus on a consolidated measure of income earned by all financial intermediaries, irrespective of whether they are classified as private equity funds, commercial banks, insurance companies, or anything else.

1.1 Raw Data

The paper uses a lot of data sources. To save space all the details regarding the construction of the series are provided in a separate online appendix. I focus on the following measure:

$$\frac{y_t^f}{y_t} = \frac{\text{Income of Finance Industry}}{\text{Aggregate Income}}.$$

Conceptually, the best measure is total income (or value added), which is the sum of profits and wages. Whenever possible, I therefore use the GDP share of the finance industry, i.e., the nominal income of the finance industry divided by the nominal GDP of the U.S. economy. One issue, however, is that before 1945 profits are not always properly measured and value added is not available. In addition, it is sometimes difficult to measure the finance industry without imputed rents from the real estate sector. As an alternative measure I then use the labor compensation share of the finance industry, i.e., the compensation of all employees of the finance industry divided by the compensation of all employees in the U.S. economy.

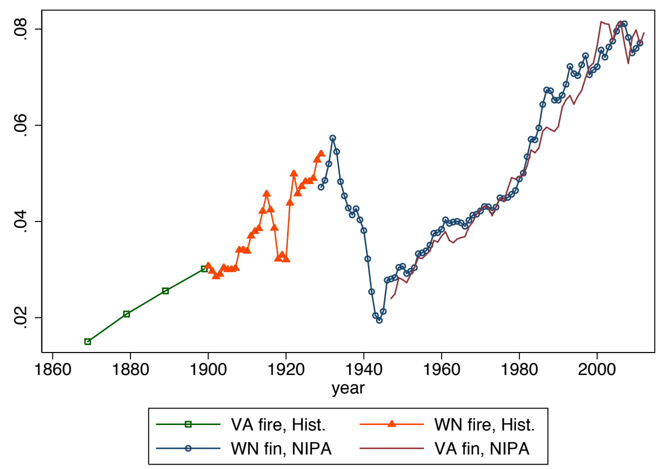
Figure 4 displays various measures of the share of the Finance and Insurance industry in the GDP of the United States estimated from 1870 to 2009. For the period 1947-2009, I use value added and compensation measures from the Annual Industry Accounts of the United States, published by the Bureau of Economic Analysis (BEA). For the

[6] See for instance Adrian and Shin (2008), Reinhart and Rogoff (2009), Krishnamurthy (2009), Acharya et al. (2009), and Scharfstein and Sunderam (2011) for recent discussions of these issues.

post-war period, the two measures display the same trends. This means that, in the long run, the labor share in the finance industry is roughly the same as the labor share in the rest of the economy (in the short run, of course, profit rates can vary). 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 (Carter et al., 2006).[7]

There are three important points to take away from Figure 4. First, the finance income share varies a lot over time. Second, the measures are qualitatively and quantitatively consistent. It is thus possible to create one long series simply by appending the older data to the newer ones. Third, finance as a share of GDP 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.

Figure 4: Income Share of Finance Industry



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.

1.2 Adjusted Measures

Before discussing theoretical interpretations it is useful to present adjusted series that take into account wars, globalization, and the rise in services.

Wars. 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 then 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

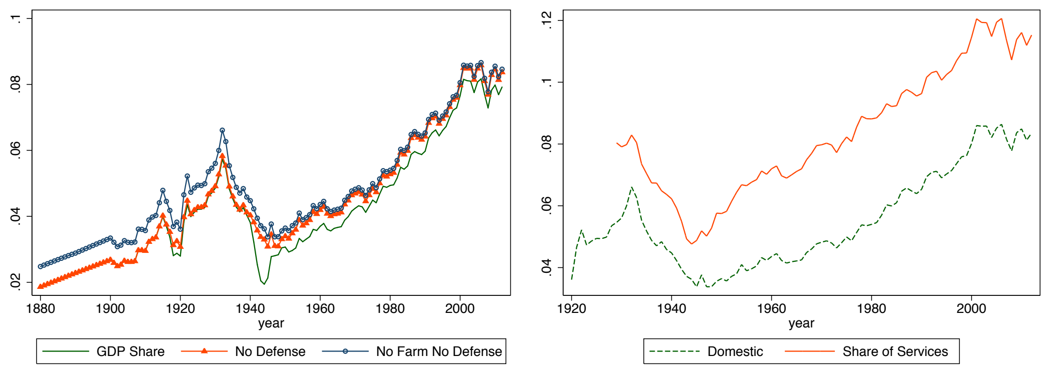
[7]Other measures based on Martin (1939) and Kuznets (1941) give similar values. More details regarding the various data sources can be found in Philippon and Reshef (2012) and in the Data Appendix

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 5 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 non-defense GDP.

Figure 5: 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 4. "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.

Other Services. Is finance different from other service industries? Yes. The right panel of Figure 5 also plots the share of finance in service GDP. It is mechanically higher than with 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. Figure 4 shows finance income divided by U.S. GDP. This might not be appropriate if financial firms export some of their services abroad. It turns out, however, that globalization does not account for the evolution of the finance income share. There are two ways to show this point.

The right panel of Figure 5 displays the ratio of *domestic* finance income to (non-defense) GDP. Domestic income is defined as income minus net exports of financial services. 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. In all case, the adjustments are small.

The timing of globalization also cannot explain the evolution of the U.S. financial sector. Estevadeordal et al. (2003) show that the period 1870-1913 marks the birth of the first era of trade globalization (measured by the ratio of trade to output) and the period 1914-1939 its end. The period between 1918 and 1930, however, is the first large scale increase in the size of the finance industry, precisely as globalization recedes. For the more recent period, Obstfeld and Taylor (2002) and Bekaert et al. (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.

2 Quantity of Intermediated Assets

2.1 Theoretical Background

To organize the discussion I use a simple model economy consisting of households, a non-financial business sector, and a financial intermediation sector. The model assumes homogenous borrowers within each group. Heterogeneity and quality adjustments are discussed in details in Section 3. This section simply highlight some important properties of the model, which is described in details in the Appendix.

Income of the finance industry. The model assumes that financial services are produced under constant returns to scale. The income of the finance industry y_t^f is given by

$$y_t^f = \psi_{c,t} b_{c,t} + \psi_{m,t} m_t + \psi_{k,t} k_t, \tag{2}$$

where $b_{c,t}$ is consumer credit outstanding, m_t are holdings of liquid assets, and k_t is the value of intermediated corporate assets. The parameters $\psi_{i,t}$'s are the unit cost of intermediation, pinned down by the intermediation technology. The model therefore says that the income of the finance industry is proportional to the quantity of intermediated assets, properly defined. The model predicts no income effect, i.e., no tendency for the finance income share to grow with per-capita GDP. This does not mean that the finance income share should be constant, since the ratio of assets to GDP can change. But it says that the income share does not grow mechanically with total factor productivity. This is consistent with the evidence presented below.[8]

[8]The fact that the finance share of GDP in Figure 4 is the same in 1925 and in 1980 makes is already clear that there is no mechanical relationship between GDP per capita and the finance income share. Similarly, Bickenbach et al. (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.

Corporate finance. As far as corporate finance is concerned, the model is fundamentally a user cost model. The key equation is

$$k^{\alpha} = \frac{1-\alpha}{r+\delta+\psi_k}, \tag{3}$$

where α is the labor share, δ the depreciation rate, and r the equilibrium interest rate. Improvements in corporate finance (a decrease in ψ_k) lead to a lower user cost of capital, a higher capital/output ratio, and a higher real wage. Despite its simplicity, this framework can account for business risk management. One well-understood benefit of risk management is to reduce the cost of financial distress, which is the net present value of the deadweight losses incurred in states of the world where firms are in financial distress. These deadweight losses include bankruptcy costs and foregone investment opportunities (see Froot et al. (1993) for a classic contribution, and Almeida and Philippon (2007) for a discussion of how to capitalize these costs). Standard models assume that financial distress can destroy a fraction of the firm's assets. From an ex-ante perspective, this is equivalent to a higher depreciation rate δ. Conversely, risk management offers one way to lower δ. This is important because there is often some confusion about how to take into account the economic value of financial derivatives. The value of risk management is capitalized in the market value of assets and the user cost $r+\delta+\psi_k$ should account properly for improvements in risk management (see the Appendix for a more formal discussion).

Household finance. A significant part of the growth of the finance industry over the past 30 years is linked to household credit. The model provides a simple way to model household finance. In the model, borrowing costs act as a tax on future income. If ψ_c is too high, no borrowing takes place and the consumer credit market collapses. The model also incorporates liquidity services provided by specific liabilities (deposits, checking accounts) issued by financial intermediaries.

In the remainder of this section, I construct empirical proxies for b, m and k, and I discuss how they can be aggregated into one measure of the quantity of intermediated assets.

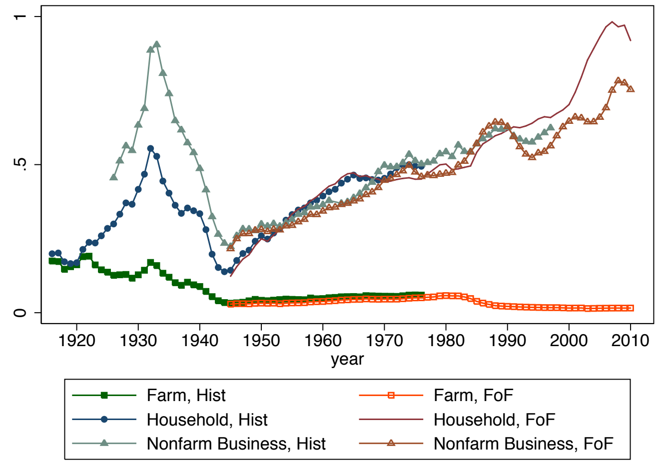
2.2 Debt and Equity

Figure 6 presents credit liabilities of farms, households and the business sector (corporate and non-corporate). 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. Two features stand out. First, the non-financial business credit market is not as deep even today as it was in the 1920s. Second, household debt has grown significantly over the post-war period.[9]

To extend the credit series before 1920, I use data on home mortgages provided by Schularick and Taylor (ming). I also use the balance sheets of financial firms. I measure assets on the balance sheets of commercial banks, mutual

[9]I have also constructed credit liabilities of financial firms. 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 6: Debt over GDP



Notes: "FoF" is flow of funds, "Hist" is Historical Statistics of the United States. Business includes non-farm corporate and non-corporate debt.

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 finance industry not only manages existing assets, but it also originates new assets and replaces old ones as they expire. It is therefore useful to consider stocks and flows separately. Figure 7 shows the issuances of corporate bonds by non-financial corporations as well as a measure of household credit flows.[10] Note that issuances collapse in the 1930s when the debt to GDP ratio peaks, in part because of deflation. There is thus a difference of timing between measures of output based on flows (issuances) versus levels (outstanding). Figure 7 also shows a measure of household debt issuance.

I use three measures of equity intermediation: total market value over GDP, IPO proceeds over GDP, and gross (non-financial) equity offerings over GDP.[11] Figure 8 shows that gross equity flows were high in the early part of the sample. The market value of equity, on the other hand, is higher in the postwar period. The IPO series will allow me to implement quality-adjustments in the next section.

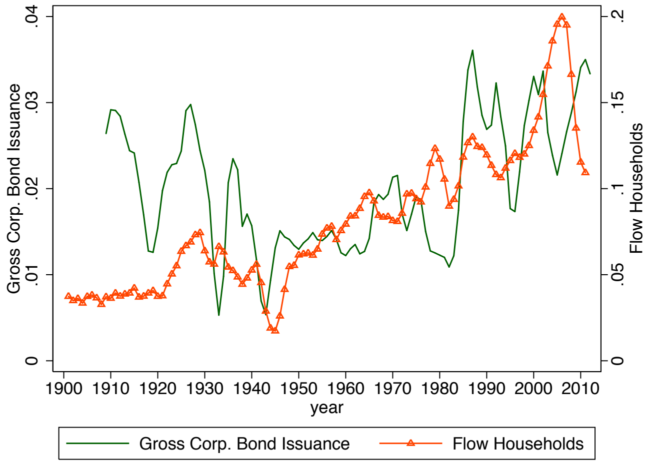
2.3 Money and Liquidity

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 total currency

[10] When I do not have a separate measure of flows, I assume a runoff rate consistent with the average ratio of flow to level, and I create the flow measure from the level series. Details are in the data appendix.

[11] Why use the *market* value of equity when thinking about intermediation? On the one hand, changes in market values may reflect changes in household risk aversion or bubbles. On the other hand, the efficiency of intermediation has a direct impact on market values. Decreases in holding and participation costs, and improvements in risk management should lead to higher market values. I assume that bubbles do not drive long term trends in equity values, and I record outstanding equity at market values. Regarding new issuances, the proper measure is clearly gross flows, not net flows.

Figure 7: Debt Flows



Notes: Gross Issuance of Corporate Bonds is a three-year centered moving average of gross issuances of bonds by non-financial firms, from Baker and Wurgler (2000). Household Issuance is based on the Flow of Funds and the Historical Statistics of the United States and

and deposits, including money market fund shares, held by households and nonprofit organizations. The left panel of Figure 9 shows the evolution of this variable.

An important element to take into account in the measurement of liquidity provision is the rise of the shadow banking system. Gorton et al. (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 et al. (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).[12]

2.4 Aggregation

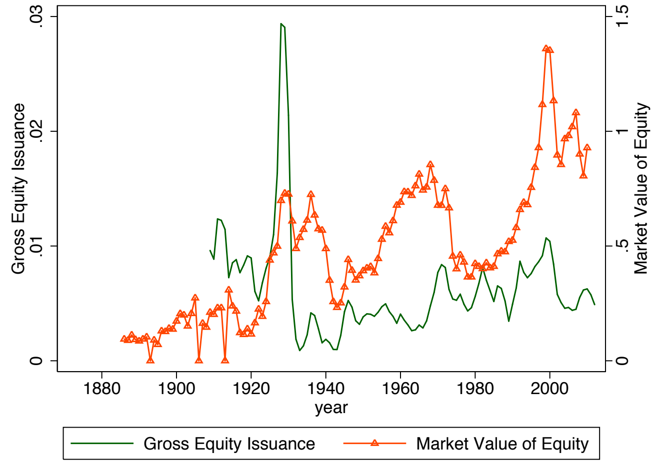
If we could observe the income flows $y_{i,t}^{f}$, associated with the three fundamental sources of revenues $i = b, m, k$, we would simply compute the unit cost as, for instance: $\psi_{c,t} = \frac{y_{c,t}^{f}}{b_{c,t}}$ where $y_{c,t}^{f}$ would be the income generated by credit intermediation for consumers. Unfortunately, there is no satisfactory way to link a particular income to a particular activity, especially over long periods of time.[13] This precludes a direct estimation of the $\psi_{i,t}$'s. We only observe the total income of the finance industry, y_t^f described in Section 1.

To make progress therefore requires new assumptions. In this Section, I assume that any change in efficiency of financial intermediation is shared proportionately among the three main activities:

[12] 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.

[13] There is an empirical problem and a conceptual problem. Empirically, our data is organized by industry (e.g., Securities, Credit Intermediation), not by function and even less by end-user. Even obtaining detailed measures of gross output is challenging. See Greenwood and Scharfstein (2013) for an enlightening discussion. But this is not only an issue of accounting. Even if we had all the data imaginable, we would still need to decide how to allocate costs among many shared activities: hedging and risk management, trading, over-head labor, etc. And 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 with the development of the originate and distribute model in banking. Therefore the problem runs even deeper if we want to make long run comparisons.

Figure 8: Equity Value and Gross Issuance over GDP



Notes: Market Value of non-financial corporate firms from the Flow of Funds and from CRSP. Gross Equity Issuance is a three-year centered moving average of gross issuances of stocks by non-financial firms, from Baker and Wurgler (2000)

Figure 9: Liquidity: Depos and Repos

Notes: Deposits, Repurchase Agreements, and Money Market Mutual Funds shares over GDP. Sources are Historical Statistics of the United States and Flow of Funds.

Assumption CRC (constant relative costs) : $\psi_{i,t} = \mu_i \psi_t$, with the normalization $\mu_c = 1$.

This assumption means that the *relative* difficulty of making a corporate loan versus a household loan remains constant over time, even though the unit costs can change. Assumption CRC will be relaxed in Section 3, albeit at the cost of much greater complexity. But for now, it allows me to defined the (weighted) quantity of intermediated assets as

$$q_t \equiv b_{c,t} + \mu_m m_t + \mu_k k_t. \tag{4}$$

Calibrating the relative costs. What remains to be done is to estimate the weights μ_m and μ_k. We can obtain indirect estimates by using the first order conditions of the model together with microeconomic evidence on the prices of various financial services. The benchmark interest rate in the economy is r. Liquid assets yield $r - \psi_m$ and (consumer) loan rates are equal to $r + \psi_c$. Direct evidence suggests that the reference rate is roughly halfway between the deposit rate and the loan rate (net of expected default). This implies $\mu_m = \mu_c$ and, given my estimated unit costs, that ψ_m is between 1.5% and 2%, which is consistent with various data sources.[14]

Corporate intermediation is made of equity and debt financing, hence $\mu_k k = \mu_k^b b_k + \mu_k^e e_k$, where e_k is corporate equity and b_k is corporate debt. Altinkilic and Hansen (2000) report fees of 3% to 4% for equity issuances and of 1% for bond issuances.[15] For flows of new issuances, I therefore assume $\mu_k^e = 3.5\mu_k^b$. These numbers seem to be in line with recent reports by large investment banks. For instance, JP Morgan's 2010 annual report suggests underwriting fees around 0.70% for debt, and around 2.46% for equity (see Appendix). Finally, as a benchmark I set $\mu_k^b = \mu_c$ so that it is equally difficult to extend credit to firms and to households.

M&As. An important activity of financial intermediaries is advising on mergers and acquisitions. Scharfstein (1988) argues that the threat of takeover can improve managerial incentives. 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. M&A fees typically range from 1% for large deals to 4% for smaller ones. I assume that merger fees are 2% of the volume. This assumption is probably a bit higher than the weighted average fee, but there are also probably some ancillary activities associated with mergers and for which the finance industry is compensated.

Flows and Stocks of Intermediated Assets. I construct two measures, one for the flow of new intermediation, one for the stock of outstanding intermediated assets.[16] Some activities are more naturally linked to flows (screening, IPO fees, etc.), some are more naturally linked to stocks (debt restructuring, asset management, etc.). The flow measure is

$$q_t^{flow} = b_{c,t}^{flow} + b_{k,t}^{flow} + 3.5e_{k,t}^{flow} + M\&A_t. \tag{5}$$

[14] The "user cost and reference rate" approach is the one used by the Bureau of Labor Statistics (BLS) to construct Producer Price indexes for Commercial Banking and Savings Institutions. The services for which indexes are available include: Loans, Deposits, Trust services, and Other banking services. To measure prices in these industries, PPI has implemented a user-cost methodology. The user cost for a financial service is the difference between the revenue it generates and the sum of its implicit and explicit costs. To measure these costs, interest is allocated between loans and deposits by means of a reference rate. Hood (2013) shows that, in 2008Q1, the loan rate (net of expected default) is 6.18%, the deposit rate is 2.74%, and the reference rate is 4.35%. This would be consistent with $r = 0.0435$, $\psi_c = 0.0183$ and $\psi_m = 0.0161$. In figure 6.4 of Fixler (2009) the reference rate is also halfway between the deposit rate and the loan rate. Table 2 in the Appendix presents other relevant rates and returns. Over the period 2002-2011, the Vanguard Short Term Treasury fund has returned 3.65%, with an expense ratio of 0.22 (Vanguard data accessed on March 11, 2012). This is the opportunity cost of cash, but cash is a relatively small fraction of liquid assets. Over the same period, the Vanguard Prime Money Market fund has returned 1.82% with an expense ratio of 0.20. The difference in returns between these two essentially risk free instruments is 1.8%. Similarly, the difference between the 1-month CD rate and the 10-year Bond rate is 1.8%, while the difference between the 10-year Bond and the conventional mortgage is 1.76%. One can also interpret ψ_m as the cost of creating liquid assets. This cost can be charged as a redemption fee for investors. This fee is around 2% and is consistent with the trading costs incurred by mutual funds upon withdrawals (see Chen et al. (2010) for a discussion).

[15] For households we can look at the mortgage market. Sirmans and Benjamin (1990) report fees of 0.50% to 0.70%. For other types of consumer credit these fees are certainly larger.

[16] I also need to take into account the debt of the government. The issue is which weight to apply. Government debt is risk-free and liquid, and it might actually help the functioning of financial markets (Krishnamurthy and Vissing-Jorgensen (2010), Greenwood et al. (2011)). But any long term debt carries duration risk, and it needs to be traded, so intermediation requirements are not exactly zero. To be conservative, I assume that government debt intermediation requirements are 1/10 of that of private debt: $\frac{\mu_g}{\mu_c} = 0.1$. The results are essentially unchanged if I set this to zero instead.

It corresponds to gross issuances of debt and equity, plus the value of mergers and acquisitions. The stock measure is simply

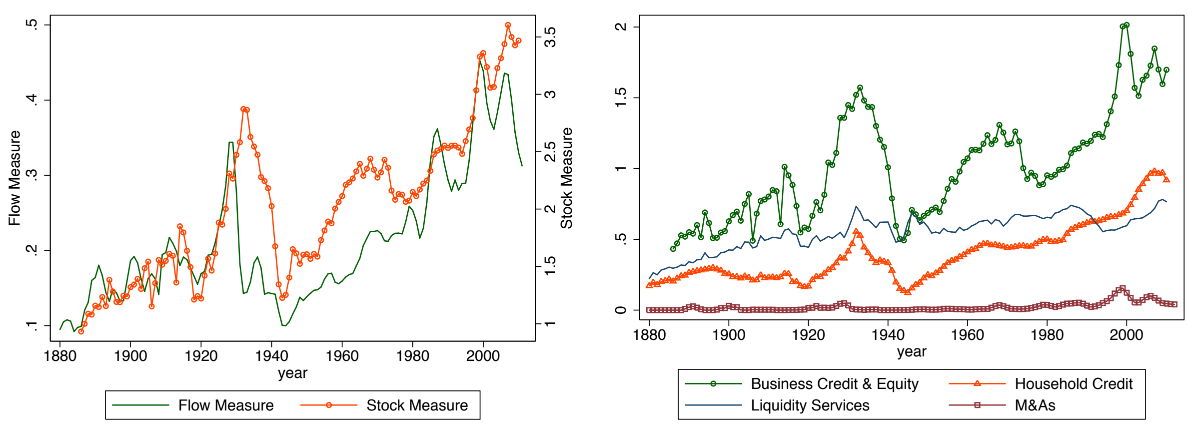
$$q_t^{level} = b_{c,t}^{level} + b_{k,t}^{level} + e_{k,t}^{level} + m_t. \tag{6}$$

Note that $e_{k,t}^{level}$ is the *market* value of equity, as discussed earlier. Note also that the liquidity measure is only a level measure, and that the M&A measure is only a flow measure. Finally, the total measure of intermediated assets is

$$q_t = q_t^{flow} + q_t^{level}. \tag{7}$$

The aggregate flow and stock measures $(q_t^{flow}, q_t^{level})$ are displayed in the left panel of Figure 10. The flow measure is an order of magnitude smaller than the stock measure. The flow measure collapses quickly during the Great Depression while the level measure peaks later and is exacerbated by deflation. A similar pattern emerges during the Great Recession. Overall, the stock measure increases more in recent years, driven by the market value of corporate equity and by the size of the household debt market.

Figure 10: Intermediated Assets over GDP, Stocks and Flows



Notes. Left panel shows aggregate measures of stocks and flows of intermediated assets in the U.S., as shares of GDP. Right panel shows the sum of stocks and flows sorted by end-user.

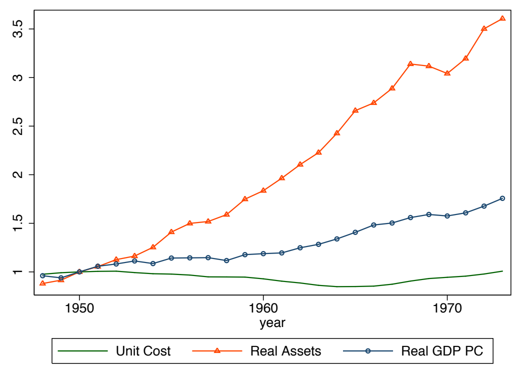
The right panel of Figure 10 presents the total measures corresponding to 4 broad functions discussed earlier: credit and equity intermediation services to firms, credit intermediation services to households, liquidity services, and M&A activities. It is clear from Figure 10 that the intermediation series for firms and households are the most volatile ones. There is also a significant increase in liquidity services in the 2000s. M&As play some role mostly in the 1990s. By construction, the sum of the two series in the left panel 10 is the same as the sum of the four series in the right panel, and is equal to the measure of intermediated assets, q, in Figure 2 in the Introduction.

2.5 Evidence of Constant Returns to Scale

Figure 3 in the Introduction shows the raw estimate of the cost of financial intermediation ψ_t, defined as income divided by intermediated assets. For income, I use domestic income, i.e., income minus net exports, as explained in Section 1. Before discussing quality adjustments in the next section, I present evidence of constant returns to scale in financial intermediation.

An important assumption of the model is that financial services are produced under constant returns to scale. Figure 11 presents evidence consistent with this assumption. It uses the period 1947-1973, for two reasons. First, the post-war data is the most reliable, and stopping in 1973 allows me to exclude major oil shocks, inflation and other factors that might create short term noise in my estimates. Second, as I will discuss shortly, quality adjustments are less important over this period than either before or after. Since these adjustments are difficult to implement, it is more convincing to first present the evidence without them.

Figure 11: Constant Returns to Scale



Notes: Unit cost of financial intermediation, real intermediated assets, and real GDP per capita. Series normalized to one in 1950.

From 1947 to 1973, real GDP per-capita increases by 80% and real financial assets by 250% (measured in constant dollars), but my estimate of the unit cost of intermediation remains fairly constant (all series are presented as ratios to their values in 1950). By 1970 people are a lot richer, financial markets are a lot larger, but the unit cost is exactly the same as in 1950. This provides clear evidence that the production of financial services has constant returns to scale.

3 Quality Adjustments

The quantities of intermediation should be adjusted for the difficulty of monitoring/screening borrowers. Without proper quality adjustment, the unit cost measure could register spurious changes in intermediation.

3.1 Corporate Finance

The homogenous borrower model described above is a useful benchmark, but it fails to capture some important features of corporate finance. To give just one example, corporate finance involves issuing commercial paper for blue chip companies as well as raising equity for high-technology start-ups. The monitoring requirements per dollar intermediated are clearly different in these two activities. Measurement problems arise when the mix of high- and low-quality borrowers changes over time. Constant heterogeneity does not pose a problem: it amounts to a simple rescaling of the unit cost in Figure 3. Changes in the share of hard-to-monitor projects, however, present a challenge.

Let us therefore consider a simple moral hazard model with heterogenous firms. If a firm hires n workers it produces $f(n)$ units of output, where f is increasing and concave. Firms choose employment to maximize (detrended) net income $\pi(w) \equiv \max_n f(n) - wn$. There are two types of firms, l and h, that differ in their cash on hand x (equivalently in their retained earnings or their pledgeable collateral). I assume that $x^l < x^h$ and I refer to l-firms as low cash firms. There is an exogenous potential supply k_h of h-firms and free entry of l-firms.[17] To capture financial intermediation in a tractable way, I assume that capital can be diverted. The Appendix describes the details of moral hazard and endogenous monitoring. The key point is that the model delivers the following monitoring demand function

$$\mu(x) = r + \delta + \varphi - \pi(w) + (1+r)(\xi - x),$$

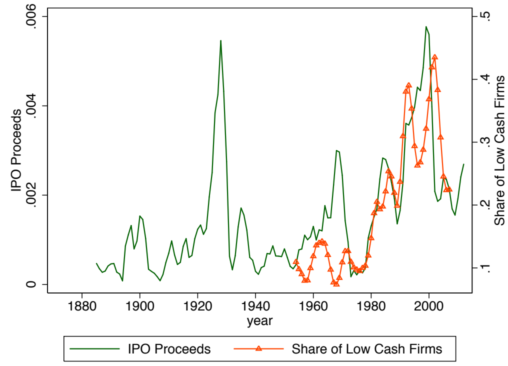
where ξ is the fraction of capital that can be diverted if there is no monitoring and φ is a proportional intermediation cost, akin to asset management fees. The function $\mu(x)$ measures the quantity of intermediation services required for a firm with cash on hand x. Firms with high values of x require less monitoring than firms with low values of x. The unit cost of monitoring is ζ_t and the income received by intermediaries for the monitoring activity is $\zeta_t \bar{\mu}_t$ where $\bar{\mu}_t \equiv \mu_h + (1+r)(x_h - x_l)s$, and $s \equiv \frac{k_l}{k_l + k_h}$ is the share of low cash firms in aggregate investment. The total income for corporate intermediation services is

$$y^f_{k,t} = \varphi_t k_t + \zeta_t \bar{\mu}_t.$$

Similarly, external finance (the quantity of monitored assets) is $\bar{b}_k = 1 - x^h + (x_h - x_l)s$. Note that the unit cost of external finance $\zeta_t \frac{\bar{\mu}_t}{\bar{b}_t}$ depends on the intensity of monitoring $\bar{\mu}/\bar{b}$, which changes with the share of low cash firms s. The parameter of interest is ζ_t which captures the true efficiency of financial intermediation. To recover ζ_t, I need to estimate $\bar{\mu}/\bar{b}$.

[17] Let k_t be the (endogenous) number of active firms, and let n_t be employment per-firm (so aggregate employment is $\bar{n}_t = k_t n_t$). The number k_h captures the extent to which investment opportunities occur in established companies. I assume that it is given by technology, and indeed, the data supports the view that large changes in k_h are driven by large scale technological change (electricity, information technology). Note that the number of low-cash firms $k_l - k_h$ is endogenous, and in particular, highly dependent on financial intermediation. So the way the model is going to interpret the 1990s is that established firms were not the ones able to promote the IT revolution. Instead it had to be younger firms, that are cash poor and therefore more dependent on financial intermediation. To the extent that we actually observe a large entry of young firms, the model will infer that financial intermediation must have been relatively efficient.

Figure 12: Inputs for Corporate Finance Quality Adjustments



Notes: The share of low cash firms is from Philippon (2008). The IPO series is a three-year centered moving average of IPO proceeds over GDP. Sources are Jovanovic and Rousseau (2001) and Ritter (2011).

Philippon (2008) uses Compustat to construct an empirical proxy for s, namely the share of aggregate investment that is done by firms that must borrow more than 3/4 of their capital spending. The measure is displayed Figure 12. Following Eisfeldt and Rampini (2006) I have also computed measures of investment that include capital reallocation by adding acquisitions minus sales of used capital for each firm. All these measures are similar and suggest that the intensity of corporate finance was higher in the 1980s and 1990s than in the 1960s. Since these measures are based on Compustat data, they are available only from 1950 onwards (at best). Figure 12 also shows IPO proceeds, based on the work of Jovanovic and Rousseau (2001) and Ritter (2011). The two series are highly correlated in the post-war period, and I use the IPO series to extend the low cash share series before 1950. As argued by Jovanovic and Rousseau (2001), the IPO market of the 1920s was remarkably active, even compared to the one of the 1990s: IPO firms were of similar ages, and the proceeds (as share of GDP) were comparable.

3.2 Household Finance

On a per-dollar basis, it is more expensive to lend to poor households than to wealthy ones, and relatively poor households have gained access to credit in recent years.[18] To capture this idea, I assume that there is a continuum households and that there is a fixed cost to borrowing κ, in addition to the marginal cost φ. Income inequality among households is captured by the labor endowment η. The model is described in details in the Appendix. The model features both an extensive margin (participation of households in the credit market) and an intensive margin (how much each household borrows). The extensive margin is characterized by the cutoff $\hat{\eta}$ such that only households with income above $\hat{\eta}$ use the credit market. The fraction of households (of a given generation) who have access to credit is therefore $1 - F(\hat{\eta})$ where F is the c.d.f of η.

[18]Using the Survey of Consumer Finances, Moore and Palumbo (2010) document that between 1989 and 2007 the fraction of households with positive debt balances increases from 72% to 77%. This increase is concentrated at 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 the mortgage market, Mayer and Pence (2008) show that subprime originations account for 15% to 20% of all HMDA originations in 2005.

The aggregate stock of household debt, relative to labor income w, is

$$\frac{\bar{b}_c}{w} = \frac{1+\gamma}{2+r+\gamma} \int_{\eta > \hat{\eta}} \left(\left(\lambda - (1-\varphi)^{-1} \right) \eta - \kappa \right) dF(\eta),$$

where γ is the rate of growth of the economy, and λ is the slope of life-cycle earnings, which determines the desire to borrow in order to smooth consumption. The income the finance industry receives from consumers credit is

$$y_c^f = \varphi \bar{b}_c + \kappa w (1 - F(\hat{\eta})).$$

3.3 Calibrated Model

The last step is to calibrate the model and construct the required quality adjustments. I rely as much as possible on micro evidence to pin down the parameters of the model. I can then reduce the number of unknown parameters to 7, which I estimate using 8 moments, so the model is slightly over-identified. An important variable is the income of the finance industry:

$$y^f = \varphi \left(k + \bar{b}_c \right) + \zeta \bar{\mu}(s) + \kappa w (1 - F(\hat{\eta})) + \psi_m m. \tag{8}$$

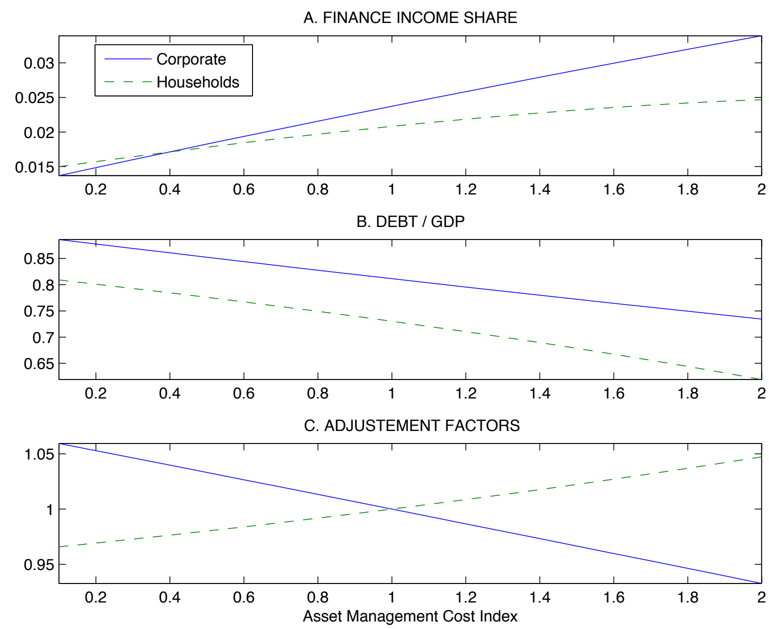
I have assumed that the linear cost (asset management) φ is the same for corporate and household finance. The parameters s and $\hat{\eta}$ capture changes in the characteristics of borrowers. I use 1989 as a reference year because of data availability. The details of the calibration are presented in the Appendix. The model matches the size of the various markets, the fraction of low cash firms, the participation rate of households in credit markets, and the income of the finance industry, all measured in 1989. The implied parameters are reasonable. For instance, I estimate a fixed borrowing cost κ of 2%. In the model, the finance industry earns 1.35% of GDP from liquidity, 2.08 from household credit, and 2.37 from business intermediation, for a total of 5.8% of GDP.

The calibrated version can then be used to understand the qualitative properties of the model and the biases that could arise in the measurement of financial intermediation. There are two types of biases. The first type of bias is that, holding intermediation technology constant, changes in the characteristics of borrowers can affect the measured unit cost of intermediation. The second type of bias comes from changes in the intermediation technology itself.

Figure 13 studies the impact of changes in the unit cost of asset management φ, which is calibrated to 1% in 1989. An increase in the cost of asset management increases the finance share of GDP (Panel A) and decreases the size of the credit market (Panel B), as expected. Note that the model, unlike the data, allows me to separate the income received from corporate finance and from household finance. The question is whether the unit costs $\psi_k = y_k^f / k$ and $\psi_c = y_c^f / \bar{b}_c$ correctly capture the changes in φ.[19] Panel C shows that the answer is "almost". Even

[19] The model gives a mapping Q from the parameters $\chi = (\varphi, \kappa, k_h, ..)$ to the equilibrium objects $\left(k, s, \bar{b}_k, \bar{b}_c, \hat{\eta}..\right) = Q(\chi)$. The income of the finance industry y^f in equation (8) depends on $Q(\chi)$ and on the intermediation technology: I write it as $y^f = Y(\chi, Q)$. In the model, this income can be further decomposed into the components coming from different types of intermediation. We have

Figure 13: Quality Adjustments for Asset Management Costs



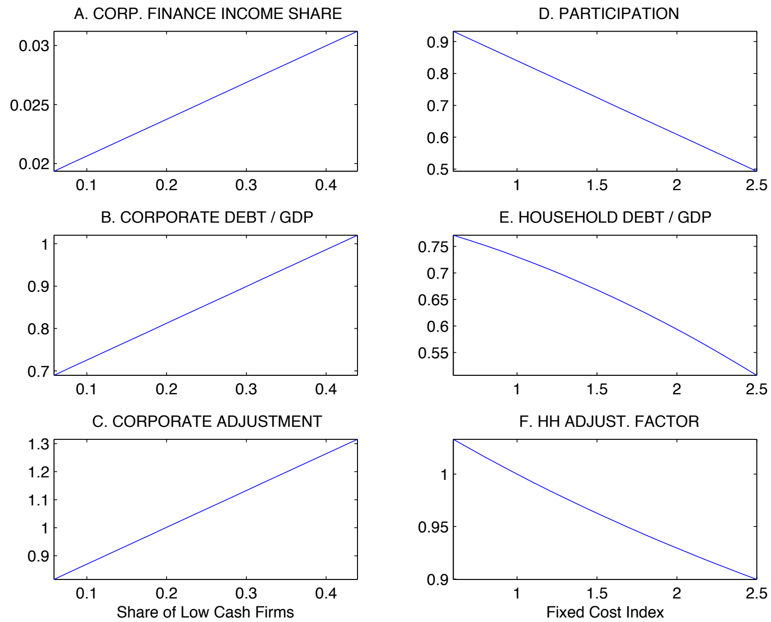
Notes: Comparative statics using the calibrated model (Table 5). The horizontal axis is φ/φ_0 where φ_0 is calibrated to 1% in 1989. Panel A presents separately the income from corporate finance services and the income from household finance services. Panel B shows firm's external finance and household debt. Panel C shows the adjustment to the particular class of intermediated assets needed to remove the bias in the measurement of intermediation costs. When the adjustment is above one, the unit cost is overestimated and the quantity of assets must be scaled up to obtain the correct estimate.

for very large changes in φ (from 0 to twice the benchmark value), the bias barely exceeds 5%. In a sense, this is not surprising because we are only changing the linear part of the model in this experiment. When φ goes down, more potential borrowers actually borrow, and each borrower borrows more. This does not create a significant bias.

Figure 14, on the other hand, studies the highly non linear part of the model, by changing the share of low cash firms and the fixed cost of participation for households. Panels A, B and C focus on changes in the composition of firms. The exogenous forcing variable is the number of cash-rich firms k_h. Note that the figure shows the response of the various variables as a function of the observed share of low cash firm s, which is itself an endogenous variable. The reason is that s is observable in the data and will be used to make the quality adjustment. The benchmark model is calibrated using a share of low cash firms of 20% (in 1989). When this share increases, monitoring costs and external finance both increase (Panels A, B). Monitoring intensity increases, and this creates a measurement

the empirical measure q from equations (5), (6) and (7) which is simply a linear combination of the elements in Q. The measured unit cost is $\psi = y^f/q$. Starting from a benchmark equilibrium (χ_0, Q_0), suppose that we change one parameter, says φ, so that $\chi = (\zeta_0, k_{h,0}, \varphi, \kappa_0, ..)$. Using the model, we can compute Q, q, and y^f, and the measured unit cost $\psi = y^f/q$. The measured change in unit cost is ψ/ψ_0. If the model is linear in the parameter of interest as in Section 2, then we have $\psi/\psi_0 = \varphi/\varphi_0$. But in general the model is not linear. To fix intuition, suppose that $\psi/\psi_0 > 1$ so we looking at a perceived increase in the unit cost. To adjust this measure, I define $\hat{y}_0^f$ as counter-factual income if we wanted to obtain Q with the initial technology ζ_0, in other words $\hat{y}_0^f = Y(\chi_0, Q)$. The conceptually correct change in the unit cost is $y^f/\hat{y}_0^f$ since by construction Q is unchanged. If $\psi/\psi_0 > y^f/\hat{y}_0^f$, then the empirical measure overestimates the change in unit cost. Then the adjustment is defined as $\psi\hat{y}_0^f/\psi_0 y^f$. This adjustment has the property that if I use it to artificially scale up q, I recover the correct value for $\psi/\psi_0 = y^f/\hat{y}_0^f$. This adjustment can be applied to the entire amount of intermediated assets, or to particular classes, such as b_c, b_k, etc.

Figure 14: Quality Adjustments for Firms Characteristics and Participation Costs



Notes: Comparative statics using the calibrated model (Table 5). Adjustment is the scaling up or down of the particular class of intermediated assets needed to remove the bias in the measurement of the unit cost. The forcing variables are k_h in panels A, B,C and κ in panels D,E, F.

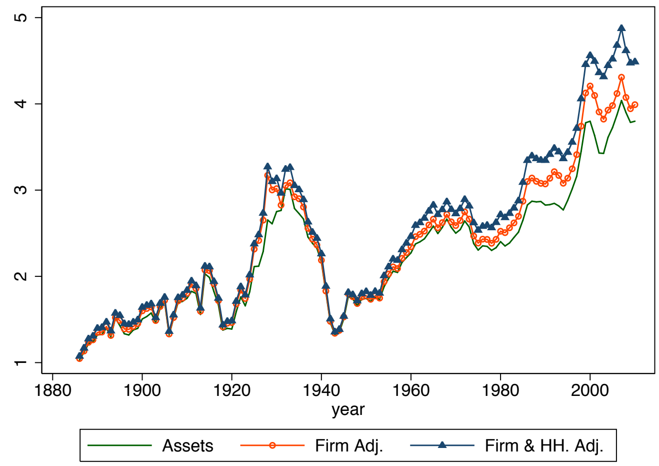
bias in the sense that the perceived unit cost increases. If the share is 40%, the model says that external corporate finance should be scaled up by roughly 25% in order to remove the induced bias in the measurement of the unit cost.

Panels D, E and F in Figure 14 focus on changes in the availability of household credit induced by exogenous changes in the fixed cost κ. When κ increases, some relatively poor households are priced out and participation in the credit market falls (D). The model is calibrated to a participation of 84% and a household debt to GDP ratio of 73% in 1989. If κ doubles, the participation rate and the debt/GDP ratio drop to approximately 60%. The participation rate drops more than the debt/GDP ratio because rich households still borrow, and they typically borrow more (when young) that poor households. These non linear composition effects create again a significant bias.

3.4 Adjusted Unit Cost

The goal of this section is to use the calibrated model presented above to adjust the asset series of Figures 10. The first step is to choose which adjustments to make. I take away from Figure 13 that changes in the proportional cost φ are unlikely to create significant biases. I therefore focus on the other parameters. At the firm level, the choice is fairly obvious: Figure 12 shows that the share of low cash firms changes a lot over time, and Panel C of Figure 14 shows that this can create large biases. The implementation is straightforward: plug in the observed value of s, read the adjustment factor in Panel C of Figure 14, and multiply the empirical series for $\bar{b}_k$ by this factor. The implied series is "Firm Adj." in Figure 15. As expected, the adjustment is quantitatively important in the 1920s and in the 1980s and 1990s, which correspond to waves of innovation driven by new technologies.

Figure 15: Quality Adjusted Intermediated Assets



Notes: Adjustments are computed using the calibrated model. "Firm Adj." takes into account changes in the fraction of low cash firms, using the series in Figure 12 and the adjustment function in Panel C of Figure 14. "Firm & HH Adj" assumes in addition that the long term growth in household finance is driven by expanding access to credit (lower κ).

Biases in the household debt market are likely to come from changes in household participation. The adjustment is difficult because I do not have a long time series for the participation rate of households. Since the goal of this section is to assess measurement biases, I will look for the maximal adjustment by assuming that changes in the household debt to GDP ratios are driven by changes in the fixed cost κ. [20] The series "Firm & HH Adj." in Figure 15 shows the output measures with quality adjustments for both corporate and household finance. Adjustments to consumer credit matters mostly after 1970. As argued earlier the quality adjustments are small between 1947 and 1973, which makes it an ideal period to test the constant returns to scale assumption. The adjustment is large in the recent part of the sample. After 1990 the unadjusted measure of business intermediation underestimates intermediation by about 25%.

Table 1 and Figure 16 are the main contributions of the paper. They bring together the historical/empirical work of Sections 1 and 2, and the theoretical/quantitative work of Section 3 (the adjustment for non-life insurance services is discussed below). Figure 16 shows the unit cost of financial intermediation, defined as income divided by adjusted intermediated assets. There are two main points to take away. The first is that the unit cost ratio is remarkably stable. Recall that we start from series – for income, debt, equity, etc – that fluctuate a lot over

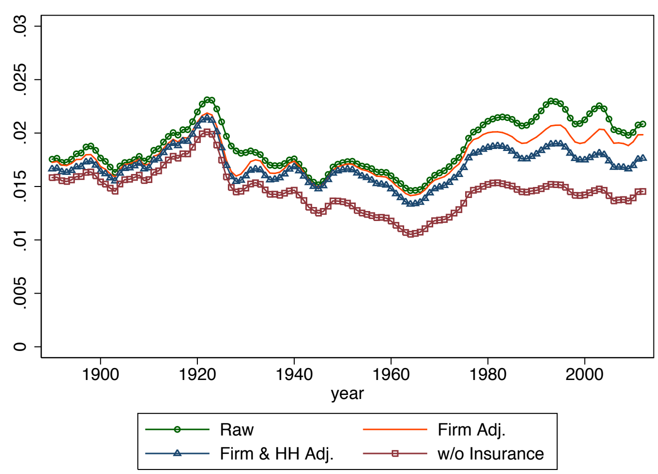
[20] There is prima facie evidence of technological change in the intermediation technology (credit scoring, etc.) that has made it easier for poor households to obtain credit. So we know that this account for *some* of the evolution of the household debt market, but we do not know precisely how much. I am going to interpret the historical time series as if the growth in consumer credit mostly reflects improvements in intermediation. I only impose the constraint that the predicted participation rate cannot exceed 100%. This constraint binds in the model in the years 2000s, which is consistent with the view that household debt growth was linked to house prices for households who already had access to credit. I have also considered the implications of changes in inequality, but I have found that these are unlikely to create significant biases. Changes in inequality typically change the debt/GDP ratio and the finance income share, but the quantitative experiments suggest that the unit cost is not severely biased.

Table 1: Estimated Financial Intermediation.

Variable	Obs	Mean	Std Dev	Min	Max
Finance Income / GDP	125	.0426	.0169	.0194	.0817
Finance Income / GDP w/o Ins	125	.0362	.0132	.0162	.0671
Intermediated Assets / GDP	125	2.245	.726	1.032	4.039
Inter. Assets, Firm Adj	125	2.368	.816	1.047	4.309
Inter. Assets, Firm & HH Adj	125	2.511	.943	1.075	4.875
Unit Cost	125	.0187	.00229	.0146	.0231
Unit Cost, Firm Adj	125	.0178	.00183	.0141	.0219
Unit Cost, Firm & HH Adj	125	.0170	.00162	.0133	.0214
Unit Cost, Adj w/o Ins.	125	.0147	.00200	.0105	.0201

time. But their ratio is stable. The simple unit cost series has a mean of 1.87% and a volatility of 23 basis points. Quality adjustments increase the volatility of the assets series but reduce the volatility of the unit cost measure, by about 25%. The adjusted series has a standard deviation of only 16 basis points. The second main point is that the financial intermediation unit cost index is higher today than in the 1960s, and about the same as it was around 1900 and in the 1920s.

Figure 16: Quality-Adjusted Unit Cost of Intermediation



Notes: Total intermediation costs divided by quality-adjusted intermediated assets.

3.5 Insurance Services

The model is designed to account for consumption smoothing that takes place via credit markets. In the model, an improvement in household finance leads to more borrowing and better consumption smoothing.[21] Insurance

[21] For instance, Gerardi et al. (2010) find that the purchase price of a household's home predicts its future income. The link is stronger after 1985, which coincides with important innovations in the mortgage market. The increase in the relationship is more pronounced for households more likely to be credit constrained. The model captures correctly measures these effects, and consumption smoothing that entails the creation of credit flows does not create a bias in my estimation. Informal risk sharing, for instance within families, would be enter neither the income side, not the asset side of my calculations, so it should not create a bias either. The overall evidence on risk sharing is mixed. Income inequality has increased dramatically in the U.S. over the past 30 years. If financial markets improve risk sharing, however, one would expect consumption inequality to increase less than income inequality. This is a controversial issue, but Aguiar and Bils (2011) find that consumption inequality closely tracks income inequality over the period 1980-2007. Therefore it seems difficult to argue that risk sharing among households has improved significantly over time.

companies, however, provide services that are not directly related to intermediation. This is potentially an issue since the income of insurance companies is counted as a cost of intermediation, while the services provided might not be well captured by standard measures of intermediated assets. I therefore attempt a (rough) adjustment by subtracting consumption expenditures on non-life insurance services (health insurance, household insurance, motor vehicle and other transportation insurance) from the total income of intermediaries. The quantitative significance of this adjustment comes from motor vehicle insurance which grows rapidly in the 1950s and is around half a percent of GDP today, and health insurance which grows linearly to reach about one percent of GDP. Whether or not these services ought to be included in financial intermediation is debatable. On the one hand, these services differ significantly from banking and traditional intermediation services. On the other hand they *are* financial services linked to the consumption of particular goods, and they certainly affect precautionary savings decisions and therefore the size of the credit market. Removing all of these expenditures is probably an over-adjustment, so the unit cost without (non-life) insurance in Figure 16 should be seen as a lower bound on the true unit cost. The new series suggests a slight downward trend in unit cost until 1970. It does not change the main point regarding the post-war sample: the unit cost is still low in the 1960s, and the discrepancy with the 2000s is at least as large as before.

3.6 Discussion of the results

Even after taking into account the various adjustments described above, the unit cost of financial intermediation appears to be higher today than it was in the 1960s, and about the same as it was around 1900.[22] This is puzzling. Advances in information technology (IT) should lower the physical transaction costs of buying, pooling and holding financial assets. Trading costs have indeed decreased (Hasbrouck, 2009), but trading volumes have increased even more, and active fund management is expensive.[23] French (2008) estimates that investors spend 0.67% of asset value trying (in vain on average, by definition) to beat the market. Similarly, Greenwood and Scharfstein (2013) show that, while mutual funds fees have dropped, high fee alternative asset managers have gained market share. The end result is that asset management unit costs have remained roughly constant. The comparison with retail and wholesale trade is instructive. In these sectors Philippon (2012b) shows that larger IT investment coincides with lower prices and lower (nominal) GDP shares. In finance, however, exactly the opposite happens: IT investment and the income share are positively related.

A potential explanation is oligopolistic competition. The idea that the finance industry is perfectly competitive, as assumed in this paper, is certainly unrealistic. But adding a constant markup of price over marginal cost would

[22]One should keep in mind that the adjustments are likely to provide lower bounds on the unit cost. Another important point is that I measure equity at market value. In equilibrium, if the cost of holding a diversified portfolio goes down, then the value of the portfolio should go up. My measure attributes the entire secular increase in the price-earnings ratio to an improvement in financial intermediation.

[23]Why do people trade so much? Financial economics does not appear to have a good explanation yet. One explanation is overconfidence, as in Odean (1998). Recent work by Glode et al. (2010) and Bolton et al. (2011) explains why some type of informed trading might be excessive. Pagnotta and Philippon (2011) present a model where trading speed can be excessive. In these models, advances in IT do not necessarily improve the efficiency of financial markets.

not change anything to the conclusions of this paper. The issue is whether market power changes over time. The historical evidence does not seem to support the naive market power explanation, however. Regulatory barriers to entry have been reduced in banking since the 1970s and yet this is when the unit cost goes up. Conversely, if there is one period where we have a strong presumption that banks had significant market power, this must be the turn of the 19th century, the late Gilded age, with the likes of John Pierpont Morgan running the major banks. But this is not a period where the unit cost of intermediation is high. The link between market power and unit cost does not seem obvious, and more research is needed on this topic.

Another plausible explanation is that my measures do not capture the social value of information production in financial markets. This effect is difficult to capture because it can show up as an improvement in TFP with little impact on the aggregate quantity of assets. The only way to test the information production hypothesis is then to estimate directly the informational content of asset prices. This is an active area of research. Bai et al. (2011) ask if current firm-level equity and bond prices predict future firm productivity and if this forecasting power has changed over time. They find 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 in that sense financial markets have become more liquid, this does not appear to translate into "better" 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)).

4 Conclusion

I have provided benchmark measures for the aggregate income of the finance industry, the quantity of intermediated assets, and the unit cost of financial intermediation. The income of the finance industry as a share of GDP fluctuates a lot over time. These fluctuations are mostly driven by equally large fluctuations in the quantity and quality of intermediated assets. The unit cost of financial intermediation represents an annual spread of 1.87% on average. The unit cost of intermediation does not seem to have decreased in recent years, despite advances in information technology, changes in the organization of the finance industry, and despite the growth of new markets, notably for financial derivatives.

Appendix

A Benchmark Model

In this section I briefly describe a model neoclassical growth model with financial intermediation.[24] This model clarifies the comparative statics and provides guidance for the aggregation of financial claims. The model economy consists of households, a non-financial business sector, and a financial intermediation sector. Long term growth is driven by labor-augmenting technological progress $A_t = (1+\gamma) A_{t-1}$. In the benchmark model borrowers are homogenous, which allows a simple characterization of equilibrium intermediation. I discuss heterogeneity and quality adjustments later.

A.1 Technology and Preferences

In the model, the finance industry provides two kinds of services to households: liquidity and credit, for transactions and consumption smoothing. Since household debt has an important life-cycle component (i.e., mortgages), I consider a setup with two types of households: some households are infinitely lived, the others belong to an overlapping generations structure.[25] Households in the model do not lend directly to one another. They lend to intermediaries, and intermediaries lend to firms and to other households.

Long-Lived Households

Long-lived households (index l) own the capital stock and have no labor endowment. Liquidity services are modeled with money in the utility function (using a cash-in-advance framework gives similar results). The households choose consumption C and holdings of liquid assets M to maximize

$$\mathbb{E} \sum_{t \geq 0} \beta^t u\left(C_t, M_t\right).$$

I specify the utility function as $u\left(C_t, M_t\right) = \frac{\left(C_t M_t^{\nu}\right)^{1-\rho}}{1-\rho}$. As argued by Lucas (2000), these homothetic preferences are consistent with the absence of trend in the ratio of real balances to income in U.S. data, and the constant relative risk aversion form is consistent with balanced growth. Let r be the interest rate received by savers. The budget constraint becomes $S_{t+1} + C_t + \psi_{m,t} M_t \leq (1+r_t) S_t$, where ψ_m is the price of liquidity services, and S are total savings.[26] The Euler equation of long lived households $u_C(t) = \beta \mathbb{E}_t \left[(1+r_{t+1}) u_C(t+1)\right]$ can then be written as

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

The liquidity demand equation $u_M(t) = \psi_{m,t} u_C(t)$ is simply

$$\psi_{m,t} M_{l,t} = \nu C_{l,t}.$$

[24] It is critical to model financial services explicitly. It is well known that the properties of two-sectors models depend on the elasticity of substitution between the two sectors (Baumol, 1967). For instance, 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. In the context of financial intermediation, I will show that the elasticity depends both on the shape of the distribution of borrowers and on the efficiency of the supply of financial services. It is therefore not possible to take this elasticity as an (exogenous) parameter.

[25] The pure infinite horizon model and the pure OLG model are both inadequate. The infinite horizon model misses the importance of life-cycle borrowing and lending. The OLG model ignores bequests, and in the simple two-periods version households do not actually borrow: the young ones save, and the old ones eat their savings. The simplest way to capture all these relevant features is the mixed model. The standard interpretation is that long-lived households have bequest motives, and are therefore equivalent to infinitely lived agents. See also Mehra et al. (2011) for a model where households save for retirement over an uncertain lifetime.

[26] See Lucas and Stokey (1987) and Sargent and Smith (2009) for a discussion of cash-in-advance models. Lucas (2000) uses the framework of Sidrauski (1967) with a more flexible functional form of the type $\left(C_t \varphi\left(\frac{M_t}{C}\right)\right)^{1-\rho}$. I use a Cobb-Douglass aggregator for simplicity given the complexity of the rest of the model. A more important difference with the classical literature on money demand is that I do not focus on inflation. Households save S at a gross return of $1+r$, while liquid assets yield $(1+r)/(1+\psi_m)$. So this model implies a constant spread between the lending rate and the rate on liquid assets. This is consistent with my interpretation of liquidity as not only money, but also money market funds shares and repurchase agreements.

Overlapping Generations

The other households live for two periods and are part of on overlapping generation structure. 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 $u\left(C_{1,t},M_{1,t}\right)+\beta u\left(C_{2,t+1},M_{2,t+1}\right)$. 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 $C_{1,t}+\psi_{m,t}M_{1,t}=\eta_1 W_{1,t}+\left(1-\psi_{c,t}\right)B_t^c$. The screening and monitoring cost is $\psi_{c,t}$ per unit of borrowing. In the second period, the household consumes $C_{2,t+1}+\psi_{m,t+1}M_{2,t+1}=\eta_2 W_{t+1}-\left(1+r_{t+1}\right)B_t^c$. The Euler equation for OLG households is

$$\left(1-\psi_{c,t}\right)M_{1,t}^{\nu(1-\rho)}C_{1,t}^{-\rho}=\beta\mathbb{E}_t\left[\left(1+r_{t+1}\right)M_{2,t+1}^{\nu(1-\rho)}C_{2,t+1}^{-\rho}\right].$$

Their liquidity demand is identical to the one of long-lived households.

Non Financial Businesses

Non-financial output is produced under constant returns technology, and for simplicity I assume that the production function is Cobb-Douglass:[27]

$$F\left(A_t n_t,K_t\right)=\left(A_t n_t\right)^{\alpha}K_t^{1-\alpha}.$$

The capital stock K_t depreciates at rate δ, is owned by the households, and must be intermediated. Let $\psi_{k,t}$ be the unit price of corporate financial intermediation. Section 3 derives the demand for intermediation services from a standard moral hazard model, but for now I take $\psi_{k,t}$ as a parameter. Non financial firms therefore solve the following program $\max_{n,K}F\left(A_t n,K\right)-\left(r_t+\delta+\psi_{k,t}\right)K-W_t n$. Capital demand equates the marginal product of capital to its user cost:

$$(1-\alpha)\left(\frac{A_t n_t}{K_t},\right)^{\alpha}=r_t+\delta+\psi_{k,t}. \tag{9}$$

Similarly, labor demand equates the marginal product of labor to the real wage:

$$\alpha\left(\frac{A_t n_t}{K_t},\right)^{\alpha-1}=\frac{W_t}{A_t}. \tag{10}$$

Financial Intermediation

Philippon (2012a) discusses in details the implications of various production functions for financial services. When financial intermediaries explicitly hire capital and labor there is a feed-back from intermediation demand onto the real wage. This issue is not central here, and I therefore assume that financial services are produced from final goods with constant marginal costs. The income share of financial intermediaries is then

$$\phi_t=\psi_{c,t}\frac{B_{c,t}}{Y_t}+\psi_{m,t}\frac{M_t}{Y_t}+\psi_{k,t}\frac{K_t}{Y_t}$$

where Y_t is aggregate GDP and $B_{c,t}$, M_t and K_t have been described above.

A.2 Equilibrium Comparative Statics

An *equilibrium* in this economy is a sequence for the various prices and quantities listed above such that households choose optimal levels of credit and liquidity, financial and non financial firms maximize profits, and the labor and capital markets clear. This implies $n_t=1$ and

$$S_t=K_{t+1}+B_t^c.$$

Let us now characterize an equilibrium with constant productivity growth in the non-financial sector (γ) and constant efficiency of intermediation (ψ). On the balanced growth path, M grows at the same rate as C. The Euler

[27] Philippon (2012a) discusses the consequences of assuming a different production function for the industrial sector. The key parameter is the elasticity of substitution between capital and labor, which is 1 under Cobb-Douglass technology. Qualitatively different results only happen for elasticity values above 6, which is far above the range of empirical estimates. Thus assuming a Cobb-Douglass technology does not entail much loss of generality.

equation for long-lived households 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 is simply pinned down by

$$\beta(1+r)=(1+\gamma)^{\theta}. \tag{11}$$

where $\theta\equiv\rho-\nu(1-\rho)$. Let lower-case letters denote de-trended variables, i.e. variables scaled by the current level of technology: for capital $k\equiv\frac{K_t}{A_t}$, for consumption of agent i $c_i\equiv\frac{C_{i,t}}{A_t}$, and for the productivity adjusted wage $w\equiv W_t/A_t$. Since $n=1$ in equilibrium, equation (9) becomes

$$k^{\alpha}=\frac{1-\alpha}{r+\delta+\psi_k}.$$

Non financial GDP is $y=k^{1-\alpha}$, and the real wage is

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

Given the interest rate in (11), the Euler equation of short lived households is simply

$$c_1=(1-\psi_c)^{\frac{1}{\theta}}c_2. \tag{12}$$

If ψ_c is 0, we have perfect consumption smoothing: $c_1=c_2$ (remember these are de-trended consumptions). In addition, all agents have the same money demand $\psi_m m_i=\nu c_i$. The budget constraints are therefore $(1+\nu)c_1=\eta_1 w+(1-\psi_c)b$ and $(1+\nu)c_2=\eta_2 w-\frac{1+r}{1+\gamma}b$. We can then use the Euler equations and budget constraints to compute the borrowing of young households

$$\frac{b_c}{w}=\frac{(1-\psi_c)^{\frac{1}{\theta}}\eta_2-\eta_1}{1-\psi_c+(1-\psi_c)^{\frac{1}{\theta}}\frac{1+r}{1+\gamma}}. \tag{13}$$

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. Household borrowing increases with the difference between current and future income, captured by $\eta_2-\eta_1$. Liquidity demand is

$$m=\frac{\nu c}{\psi_m}.$$

and aggregate consumption is

$$c=\frac{1}{1+\nu}\left(w-\psi_c b_c+(r-\gamma)k\right). \tag{14}$$

The following proposition summarizes the predictions of the theory.

Proposition 1. *The equilibrium has the following features*

(i) Along a balanced growth path with constant intermediation technology, constant demographics, and constant firms' characteristics, the finance share of GDP and the financial ratios m/y, b^c/y and k/y are constant.

(ii) Improvements in corporate finance increase y, w, k/y, c/y and m/y, but leave b^c/y constant;

(iii) Improvements in household finance increase b^c/y, c/y and m/y, but do not affect k;

(iv) Increases in the demand for intermediation increase the finance income share ϕ while supply shifts have an ambiguous impact.

Proof. Point (i): it is clear from the equations of the model. Point (ii): when ψ_k goes down, k, y and w go up, k/y goes up because of decreasing returns to capital, but equation (13) shows that b_c/y remain constant since $w=\alpha y$. From (14), c/y increases because k/y increases, and so does m/y. Point (iii): when ψ_c goes down, b_c, c and m go up. Since the user cost of capital is not affected, k and w are constant. Point (iv): suppose η_2 goes up, and $\eta_1=1-\eta_2$ goes down. From (13), b_c goes up while k, w and y are unchanged. Then c and m go down, but one can check that $\psi_c b_c+\psi_m m$ goes up and therefore ϕ increase. On the other hand, suppose now that ψ goes down. There are two effects: financial services are cheaper which pushes y^f down, but more services will be provided in equilibrium. So the net impact on y^f is theoretically ambiguous. ∎

That homogeneity in production is required for balanced growth is not surprising. What is more interesting is that it is sufficient even if the production technologies differ between the financial and non-financial sectors (see Acemoglu and Guerrieri (2008) and Philippon (2012a) for detailed discussions). Regarding liquidity demand, balanced growth comes from the assumed preferences, as discussed in Lucas (2000).

A.3 Risk management

The market for financial derivatives is extremely large. Since these contracts are in zero net supply, however, they do not enter *directly* into my calculations of intermediated assets. The question is: Should they? The answer is essentially no, because the benefits of derivatives are already *indirectly* taken into account.

One thing is clear: it would make no sense to count derivatives at face value. Rather, one should take the perspective of standard economic theory and recognize that derivatives can add real value in one of two ways: (i) risk sharing; (ii) price discovery. Bai et al. (2011) discuss price discovery so let me focus on how risk management affects my measures. Risk management among banks lowers intermediation costs as banks lay off excess risk inventories when making markets. This type of risk sharing among financial intermediaries does not create any bias in my measurements. Improvements in risk management would simply lead to lower borrowing costs and cheaper financing, and this would be correctly captured by the model.[28]

Risk management by the non-financial sector is more subtle, but the user cost framework is still the right approach. To see why, decompose the depreciation rate into an exogenous physical depreciation δ_p and an expected cost of distress δ_f. We can then endogenize the expected cost of distress by assuming that the technological production frontier of the finance industry is $\delta_f^2 + \psi^2 \geq \chi^2$, where ψ is the cost of risk management and χ indexes the production productivity frontier of risk management services. This captures the idea that it is possible to decrease financial distress by spending more on risk management. The optimal choice is simply to minimize the user cost of capital, i.e.,

$$\min \delta_f + \psi \text{ s.t. } \delta_f^2 + \psi^2 \geq \chi^2.$$

This then leads to $\delta_f = \psi = \chi/\sqrt{2}$. Then the total financial cost is $\delta_f + \psi = \chi\sqrt{2}$, and in steady state we have

$$k = \left(\frac{1-\alpha}{r + \delta_p + \chi\sqrt{2}} \right)^{\frac{1}{\alpha}}.$$

Improvements in risk management increase k, just like generic improvements in intermediation. The finance industry earns $\psi k = \frac{\chi}{\sqrt{2}} k$, so my measure of unit cost would simply recover the evolution of χ.

In the neoclassical growth model with homogenous firms, there is no value of the firm as a going concern. In the next section, I consider an extension with heterogenous firms and decreasing returns at the firm level, where the value of some firms exceeds the replacement cost of their capital. In general we can think of firm value (per unit of capital) as

$$V = \pi + \frac{1 - \delta_p - \delta_f - \psi}{1+r} V'$$

where V is the current value, V' the future value, π is the profit rate, δ_f the cost of financial distress and ψ the cost of risk management. Using the same intermediation technology as above, the optimal risk management yields $\delta_f = \psi = \chi/\sqrt{2}$. Solving for V in steady state, we have

$$V = \frac{(1+r)\,\pi}{r + \delta_p + \chi\sqrt{2}}$$

The important point here is that improvement in risk management would be capitalized in the market value of the firm. This provides a clear argument for using the market value of equity in the calculations. A caveat here is that I treat risk management and price discovery as two separate issues, but they need not be. In DeMarzo and Duffie (1991) for instance, financial hedging is fundamentally linked to private information about firm value, and

[28] 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 equivalent. Now 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 earlier. 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.

Table 2: U.S. Interest Rates and Returns, 2002-2011

Variable	Value (%)
Average 3M Treasury Bills	1.79
Average 1M Certificate of Deposits	2.14
Average 1Y Gov. Bond	2.10
Average 10Y Gov. Bond	3.95
Average Aaa Corporate	5.47
Average Prime Bank Loan	5.02
Average 30Y Conventional Mortgage	5.71
Average Baa Corporate	6.64
Vanguard Prime Money Market Fund Return	1.82
Vanguard Short Term Treasury Fund Return	3.65

Source: FRED, and Vanguard. All values are average over 2002-2011.

in DeMarzo and Duffie (1995) hedging interacts with incentives and accounting disclosure. Those complex and fascinating issues are beyond the scope of this paper.

A.4 Calibration of basic model

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 \$26b, 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 governments and non-profits. The bank advised 311 announced M&A (a 16% market share). The bank also loaned or arranged \$350b.

B Quality Adjustments

This section presents the model with heterogenous firms and households. It extends the model of Appendix A.

B.1 Firms

This section describes what happens within a period, so I suppress the time index (remember that in the OLG setup, one period represents many years). There are k firms. Firm i is endowed with $x_i A$ and needs to borrow $(1-x_i)A$ to operate a technology that produces according to $f(n) = An^{\alpha}$. With Cobb-Douglass technology, we get net income

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

and labor demand $n = \left(\frac{\alpha}{w}\right)^{\frac{1}{1-\alpha}}$.

There are two frictions in capital markets. The first is a proportional cost φ as in Appendix A, which capture holding costs and asset management fees. The second friction comes from moral hazard. Firms' owners can divert resources and monitoring by intermediaries is used to reduce the risk of diversion. If the firm behaves well, it pays back its outside investors $(1+r)(1-x)$ and insiders receive

$$\pi(w) + 1 - \delta - \varphi - (1-x)(1+r) - \zeta\mu = \pi(w) - \delta - \varphi - r - \zeta\mu_k + (1+r)x,$$

where ζ is the unit cost of monitoring and μ the quantity of monitoring used by the firm. If the firm cheats, outside investors receive nothing and inside investors keep $(1+r)\,\xi-\zeta\mu-\mu$, where ξ measures the degree of diversion.[29] The incentive constraint is therefore $\pi(w)-\delta-r+(1+r)\,x\geq(1+r)\,\xi-\mu$. The firm seeks to minimize the cost of monitoring subject to a break-even constraint and an incentive constraint:

$$
\begin{aligned}
\min_{\mu\geq 0}\zeta\mu \qquad & s.t. \\
\pi(w) \;\geq\; & r+\delta+\varphi+\zeta\mu, \\
\pi(w) \;\geq\; & r+\delta+\varphi+(1+r)(\xi-x)-\mu.
\end{aligned}
$$

The first constraint binds for marginal firms, i.e., firms that are indifferent between entering and staying out, and all firms must satisfy their incentive constraints. This defines the required level of monitoring as a function of aggregate parameters (r, π) and of the firm's level of free cash flows x:

$$\mu(x)=\max\left(0;r+\delta+\varphi-\pi(w)+(1+r)(\xi-x)\right). \tag{15}$$

We consider a model with two types of firms, $i=l,h$ with $x_l<x_h$. We study equilibria where the number of high cash ventures is (exogenously) given by k_h and where l is the marginal type (i.e., k_h is not to large enough to exhaust the investments needs of the economy). There is free entry of low types, therefore we have

$$\pi(w)=r+\delta+\varphi+\zeta\mu(x_l),$$

and substituting in the IC constraint we get

$$\mu(x_l)=\frac{1+r}{1+\zeta}(\xi-x_l).$$

This pins down the required profit rate $\pi(w)$, and therefore the equilibrium wage:

$$\pi(w)=r+\delta+\varphi+\frac{\zeta}{1+\zeta}(1+r)(\xi-x_l).$$

The h-firms earns rents since $\mu(x_h)<\mu(x_l)$. From (15) and assuming that $\mu_h>0$, the difference in monitoring requirements is

$$\mu_l-\mu_h=(1+r)(x_h-x_l).$$

The quantity of monitoring produced by financial intermediaries is given by

$$\bar{\mu}=\sum_{j=l,h}k_j\mu_j.$$

The income of the finance industry is made of direct intermediation income (monitoring, screening) and asset management fees

$$y^f_{k,t}=\varphi_t k_t+\zeta_t\bar{\mu}_t.$$

Focusing on aggregate borrowing, we have $\bar{b}=k_h(1-x_h)+k_l(1-x_l)$, so the intensity of monitoring, $\bar{\mu}/\bar{b}$ is

$$\frac{\bar{\mu}}{\bar{b}}(s)=\frac{\mu_h+(1+r)(x_h-x_l)\,s}{1-x^h+(x_h-x_l)\,s}$$

where $s\equiv\frac{k_l}{k_l+k_h}$ is the fraction of high monitoring (low-cash) firms in aggregate investment. The ratio of income over intermediated assets is for the monitoring activity is

$$\zeta_t\frac{\bar{\mu}_t}{\bar{b}_t}$$

[29] This formula assumes that firms cannot divert bankers' fees. The analysis is essentially the same if they can, one must simply carry an extra term $\zeta\mu_k$ in the formulas.

The parameter of interest is ζ_t which captures the true efficiency of financial intermediation. In the data I measure $\frac{\zeta_t \bar{\mu}_t}{b_t}$. To recover ζ_t I therefore need to estimate $\frac{\bar{\mu}}{b}(s_t)$.

B.2 Households

I consider the case of log preferences $\rho = 1$. Long-lived households are the same as before, therefore

$$\beta (1 + r) = 1 + \gamma$$

Short-lived households have a desire to borrow. There is a fixed cost to borrowing, in addition to the marginal cost φ. Households differ in their incomes. As before, they value their stream of consumptions as

$$V_t = \log C_{1,t} + \nu \log M_{1,t} + \beta \log C_{2,t+1} + \beta\nu \log M_{2,t+1}.$$

Money demand is unchanged, so we still have

$$\psi_{m,t} M_{i,t} = \nu C_{i,t}.$$

Normalizing by aggregate productivity A_t, we see that

$$V_t = \log c_{1,t} + \beta \log c_{2,t+1} + \Xi_t,$$

where $\Xi_t \equiv (1 + \nu)(\log A_t + \beta \log A_{t+1}) + \nu \log \frac{\nu}{\psi_{m,t}} + \nu\beta \log \frac{\nu}{\psi_{m,t+1}}$ only depends on aggregate quantities. The households' choices therefore really depend upon $\log(c_1) + \beta \log(c_2)$, assuming as before constant TFP growth.

Autarky If the household stays out of the credit market, its budget constraints are simply$(1 + \nu) c_1 = \eta_1 w$ and $(1 + \nu) c_2 = \eta_2 w$. The value of autarky is therefore

$$V_{aut} = \log(\eta_1) + \beta \log(\eta_2) + (1 + \beta) \log \frac{w}{1 + \nu}$$

Credit If the household enters the credit market, it pays the fixed cost κw at time 2. The budget constraints are $(1 + \nu) c_1 = \eta_1 w + (1 - \varphi) b$ and $(1 + \nu) c_2 = \eta_2 w - \kappa w - \frac{1+r}{1+\gamma} b$, and its Euler equation is therefore

$$c_1 = (1 - \varphi) c_2.$$

Obviously, if $\psi = 0$, we have $c_1 = c_2$. From these equation we can compute the amount borrowed as

$$\frac{b_c}{w} = \frac{\eta_2 - \kappa - \frac{\eta_1}{1-\varphi}}{1 + \frac{1+r}{1+\gamma}}. \tag{16}$$

Clearly, borrowing requires $\eta_2 > \kappa + \frac{\eta_1}{1-\varphi}$. The maximum amount of borrowing, if $\varphi = \kappa = 0$ is $\frac{b_c}{w} = \frac{\eta_2 - \eta_1}{1+\frac{1+r}{1+\gamma}}$ The value of entering the credit market is

$$V_{bor} = (1 + \beta) \log(c_1) - \beta \log(1 - \varphi).$$

We can compute the consumption levels from the budget constraints:

$$c_1 = \frac{w}{1 + \nu} \frac{(1 + r)\eta_1 + (1 - \varphi)(1 + \gamma)(\eta_2 - \kappa)}{2 + r + \gamma}.$$

Households' choices I parameterize the model in the following way. Let η be a random variable with mean of $\frac{1}{1+\lambda}$ drawn at birth. $\eta_1 = \eta$ and $\eta_2 = \lambda\eta$ with $\lambda > 1$. The choice of entering the credit market or not depends on

$$\Delta \equiv V_{bor} - V_{aut}$$

Substituting the above expressions, we get

$$\Delta = (1+\beta)\log\left(\frac{1+r+(1-\varphi)(1+\gamma)\left(\lambda-\frac{\kappa}{\eta}\right)}{2+r+\gamma}\right) - \beta\log\lambda(1-\varphi)$$

We see that

$$\Delta > 0 \Longleftrightarrow \eta > \hat{\eta}$$

where the cutoff $\hat{\eta}$ solves

$$\lambda - \frac{\kappa}{\hat{\eta}} = \frac{(2+r+\gamma)(\lambda(1-\varphi))^{\frac{\beta}{1+\beta}} - (1+r)}{(1-\varphi)(1+\gamma)}.$$

Aggregate outcome Liquidity demand is

$$m = \frac{\nu c}{\psi_m}.$$

and aggregate consumption is

$$c = \frac{1}{1+\nu}\left(w - \varphi\bar{b}_c - \kappa w(1-F(\hat{\eta})) + (r-\gamma)k\right).$$

where aggregate household debt is

$$\frac{\bar{b}_c}{w} = \frac{1+\gamma}{2+r+\gamma}\int_{\eta>\hat{\eta}}\left(\left(\lambda-(1-\varphi)^{-1}\right)\eta-\kappa\right)dF(\eta).$$

Finance income from consumer credit is

$$\varphi\bar{b}_c + \kappa w(1-F(\hat{\eta})).$$

B.3 Calibration and Mapping from the Model to the Data

Table 5 presents the parameters. Some parameters are set using values that are standard in the literature. The finance-specific parameters are chosen to match micro-data and a set of moments. The moments are presented in the second row of the table, and the implied parameters in the last row. I calibrate the model using 1989 as a reference year because the Survey of Consumer Finance is available for that year.

Table 3: Standard Parameters

Rate	Deprec.	Growth	Labor Sh.	CRRA
$r = 0.05$	$\delta = 0.1$	$\gamma = 0.02$	$\alpha = 0.7$	$\rho = 1$

Corporate Finance In the model presented here, there are two types of intermediation. Internal funds cover a fraction x_i of the capital needs, with $x_h > x_l$. The remaining $1 - x_i$ is external, and I refer to it as external finance. The management fee φ applies to the entire stock of capital while the monitoring cost applies only to $b_k = 1 - x$. To map these quantities with the data, I make the traditional assumption that external finance is debt and new issuances of debt and equity, while outstanding equity is internal. I therefore define external finance as total business intermediation minus the value of existing equity, or equivalently as the stock of debt plus issuances of new debt and equity. In 1989, the stock of existing firm debt is 0.665 GDP. Taking into account the flows of new debt and equity, monitored finance is 0.806 GDP. This is the target value for $\bar{b}_k/y$ in the model.

Following Greenwood and Scharfstein (2013), I set the asset management fee to $\varphi = 1\%$ and I apply this fee with the appropriate accounting corrections.[30] In the data, low cash firms cover about 10% of their expenditures, so I set $x_l = 0.1$. They account for 20% of investment in 1980 (see Figure 12), so I use $s_k = 0.20$ as a target. The average leverage of non financial firms (debt over debt plus equity) is 0.4, as shown for instance in Graham et al. (2012). Together with x_l and s_k, this fixes the value of $x_h = 0.62$.[31]

Beyond, x_l and x_h, the parameters governing corporate finance are the degree of moral hazard ξ and the cost of monitoring ζ. Since ζ and ξ are not really separately identified, I set a value for ξ. The restriction $\xi > x_h$ ensures that monitoring demand is positive for all firms and of course we must have $\xi < 1$ since it is measures a fraction, so I can pick any $\xi \in (0.62, 1)$. I set $\xi = 0.8$. The results do not depend on this choice. What is important, however, is the value of $\frac{\zeta}{1+\zeta}\xi$. To identify ζ, I use a target for the unit cost of external finance $\varphi + \zeta\bar{\mu}/\bar{b}_k$ of 2.05%, based on the following calculation. Bonds represent about half of credit market instruments for non financial firms in 1989. For bonds, the cost is the asset management fee of 1% plus a liquidity premium around 0.5% which in equilibrium is paid to market makers (see Almeida and Philippon (2007) for a discussion) plus a 1% issuance fee paid every 10 year (the average bond maturity in 1989). This gives 1.6%. For the remaining half, I assume a cost of 2.5%, consistent with the spread between the reference rate and the lending rate of banks, as presented in Fixler (2009). Note that 2.5% is also the management fee for alternative asset managers, which presumably should be counted as monitored finance. My target for $\varphi + \zeta\bar{\mu}/\bar{b}_k$ is therefore $(1.6+2.5)/2 = 2.05\%$.

To summarize the corporate finance calibration, I have used micro data to pin down all the parameters, except for two: the number of high-cash firms k_h, and the monitoring efficiency ζ. To make sure the model if over-identified, I attempt to match three moments: the share of low cash firms in aggregate investment $s_k = 0.20$, the unit cost of monitored finance $\varphi + \zeta\bar{\mu}/\bar{b}_k = 2.05\%$, and the quantity of external finance over GDP $\bar{b}_k/y = 0.806$.

Table 4: Parameters Estimated from Micro Data

High cash	Low cash	Asset Mgt Fee
$x_h = 0.62$	$x_l = 0.1$	$\varphi = 0.01$

Household Finance In the model there is a mass 1 of young workers and a mass 1 of old workers. Aggregate labor endowment is normalized to 1, so young workers are endowed with $\frac{1}{1+\lambda}$ units of labor on average and old workers are endowed $\frac{\lambda}{1+\lambda}$ units of labor on average, and λ drives the desire of young workers to borrow.

I assume that heterogeneity η is uniform over $\left[\frac{1-H}{1+\lambda}, \frac{1+H}{1+\lambda}\right]$ for some $H > 0$. The density inside the interval is $f(\eta) = \frac{1+\lambda}{2H}$. Old workers have the same distribution, but scaled by λ. The top earners in this economy are the old workers with high values of η. More precisely, the top 40% of old people represent the top 20% of the population (this is true for the value of λ that I consider below) and have an average income of $\frac{\lambda}{1+\lambda}\left(1 + \frac{3}{5}H\right)$. Their combined labor income is $0.4\frac{\lambda}{1+\lambda}\left(1 + \frac{3}{5}H\right)w$ relative to a total income of w. In the data the top quintile earns 46.8% of aggregate income in 1989 (Jones and Weinberg, 2000), but some of their income is capital income. In the model, the labor income share is α and the capital income share is $1 - \alpha$. I calibrate the model so that they earn 60% of capital income.[32] Therefore, H must be such that

$$0.4\frac{\lambda}{1+\lambda}\left(1 + \frac{3}{5}H\right)\alpha + 0.6(1-\alpha) = 0.468. \tag{17}$$

This pins down H given λ and α.

I also target the participation rate of households in the credit market, and the size of the household debt market. According to Moore and Palumbo (2010), in 1989, 84% of heads of household of age 45 or less had

[30] For firms, I take into account that k/y in the model is not the same as k/y in the data because of the fluctuation in the ratio of market to book equity and because my empirical measure does not capture privately held capital. The model implies k/y of 1.7 while firm intermediation in 1989 is around 1.35. To get the correct measure of income, I therefore use $\varphi_k = \frac{1.3}{1.7} \times 1.3\%$.

[31] External finance per unit of asset is $s_k(1 - x_l) + (1 - s_k)(1 - x_h)$. As explained, monitored finance is the stock of debt plus issuances of new debt and equity. To be internally consistent, I therefore set x_h such that $s_k(1 - x_l) + (1 - s_k)(1 - x_h) = \frac{0.806}{0.665} \times 0.4 = 0.485$. This gives $x_h = 0.619$.

[32] According to Saez's data, when the top decile earns 40% of total income, it earns 31.5% of labor income. If θ is the fraction of capital income earned by this group, we have $0.315\alpha + (1-\alpha)\theta = 0.4$ which implies that they earn therefore 60% of capital income. I therefore require that

positive debt balance. I therefore $1 - F(\hat{\eta}) = 0.84$, which pins down $\hat{\eta}$ given H. The size of the household debt market is $\frac{\bar{b}_c}{w} = \frac{1+\gamma}{2+r+\gamma} \int_{\eta>\hat{\eta}} \left(\left(\lambda - (1-\varphi)^{-1} \right) \eta - \kappa \right) dF(\eta)$. With a uniform distribution, we have $\int_{\eta>\hat{\eta}} \left(\left(\lambda - \frac{1}{1-\varphi} \right) \eta - \kappa \right) dF(\eta) = \frac{1+\lambda}{2H} (\bar{\eta} - \hat{\eta}) \left(\left(\lambda - \frac{1}{1-\varphi} \right) \frac{\bar{\eta}+\hat{\eta}}{2} - \kappa \right)$ with $\bar{\eta} \equiv \frac{1+H}{1+\lambda}$. Aggregate household debt is therefore $\bar{b}_c = \frac{1+\gamma}{2+r+\gamma} \frac{1+\lambda}{2H} (\bar{\eta} - \hat{\eta}) \left(\left(\lambda - \frac{1}{1-\varphi} \right) \frac{\bar{\eta}+\hat{\eta}}{2} - \kappa \right) w$. In the data, $\bar{b}_c/y$ is 0.73.

The model is over-identified by one parameter but the fit is good. The implied parameters are all reasonable. For instance, $\kappa = 2.3\%$ means that the fixed cost is 4.3% of the average wage of young workers, and 10.7% of the wage of the marginal worker (the one earning $\hat{\eta}$ who is indifferent between participating or not participating). In the model, the finance industry earns 1.35% of GDP from liquidity, 2.08 from household credit, and 2.37 from business intermediation, for a total of 5.8% of GDP.

Table 5: Estimation

Moments

	Business Debt $\bar{b}_k/y$	Unit Cost $\varphi + \zeta \frac{\bar{\mu}_k}{\bar{b}_k}$	l-Firms s	HH Debt $\bar{b}_c/y$	HH Particip. $1 - F(\hat{\eta})$	Top 20% HH Eq (17)	Liquidity m/y	Fin. Share y^f/y
data	0.806	0.0205	0.20	0.73	0.84	0.468	0.71	0.0585
model	0.811	0.0208	0.199	0.73	0.84	0.468	0.71	0.0580

Implied Parameters

Monit. Cost	H-Cash Firms	Earn. Slope	Ineq.	Fix. Cost	Liq. Demand	Liq. Supply
$\zeta = r/3.35$	$k_h/k^* = 0.622$	$\lambda = 2.07$	$H = 0.875$	$\kappa = 0.023$	$\nu = 0.0181$	$\psi_m = 0.019$

Notes: Calibration of model using data from 1989. k^* refers to the first best level of capital (obtained by setting all intermediation costs to zero).

References

Acemoglu, D. and V. Guerrieri (2008). Capital deepening and non-balanced economic growth. *Journal of Political Economy 116*(3), 467–498.

Acharya, V. V., L. H. Pedersen, T. Philippon, and M. Richardson (2009). Measuring systemic risk. Working Paper NYU.

Admati, A. R., P. M. DeMarzo, M. Hellwig, and P. Pfleiderer (2011). Fallacies, irrelevant facts, and myths in the discussion of capital regulation: Why bank equity is not expensive. Working Paper Stanford University.

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.

Almeida, H. and T. Philippon (2007). The risk-adjusted cost of financial distress. *Journal of Finance 62*(6), 2557–2586.

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 markets become more informative? Working Paper NYU.

Baker, M. and J. Wurgler (2000, October). 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.

Bazot, G. (2013). Financial consumption and the cost of finance: Measuring financial efficiency in europe (1950-2007). Working Paper Paris School of Economics.

Beck, T., A. Demirguc-Kunt, and R. Levine (2011). *Financial Structure and Economic Growth: A Cross-Country Comparison of Banks Markets, and Development*, Chapter The Financial Structure Database, pp. 17–80. Cambridge: MIT Press.

Bekaert, G., C. R. Harvey, and R. Lumsdaine (2002). Dating the integration of world capital markets. *Journal of Financial Economics 65*, 203–249.

Bernanke, B., M. Gertler, and S. Gilchrist (1999). The financial accelerator in a quantitative business cycle framework. In J. B. Taylor and M. Woodford (Eds.), *Handbook of Macroeconomics*, Volume 1C. Elsevier Science, North Holland.

Bickenbach, F., E. Bode, D. Dohse, A. Hanley, and R. Schweickert (2009, October). 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.

Buera, F. J., J. P. Kaboski, and Y. Shin (2011). Finance and development: A tale of two sectors. *The American Economic Review 101*(5), 1964–2002.

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.

Chen, Q., I. Goldstein, and W. Jiang (2010). Payoff complementarities and financial fragility: Evidence from mutual fund outflows. *Journal of Financial Economics 97*(2), 239 – 262.

Christiano, L. and D. Ikeda (2011). Government policy, credit markets and economic activity. mimeo Northwestern.

Corsetti, G., K. Kuester, A. Meier, and G. J. Müller (2011). Sovereign risk and the effects of fiscal retrenchment in deep recessions.

Curdia, V. and M. Woodford (2009). Conventional and unconventional monetary policy. mimeo Columbia.

DeMarzo, P. M. and D. Duffie (1991, April). Corporate financial hedging with proprietary information. *Journal of Economic Theory 53*(2), 261–286.

DeMarzo, P. M. and D. Duffie (1995). Corporate incentives for hedging and hedge accounting. *Review of Financial Studies 8*(3), 743–71.

Diamond, D. W. (1984). Financial intermediation and delegated monitoring. *Review of Economic Studies 51*, 393–414.

Diamond, D. W. and P. H. Dybvig (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy 91*, 401–419.

Diamond, D. W. and R. G. Rajan (2001). Liquidity risk, liquidity creation and financial fragility: A theory of banking. *Journal of Political Economy 109*, 287–327.

Eisfeldt, A. L. and A. A. Rampini (2006). Capital reallocation and liquidity. *Journal of Monetary Economics 53*, 369–399.

Estevadeordal, A., B. Frantz, and A. M. Taylor (2003, 05). The rise and fall of world trade, 1870-1939. *The Quarterly Journal of Economics 118*(2), 359–407.

Fixler, D. (2009). *Price Index Concepts and Measurement*, Chapter Incorporating Financial Services in a Consumer Price Index. University of Chicago Press.

French, K. R. (2008). Presidential address: The cost of active investing. *The Journal of Finance 63*(4), 1537–1573.

Froot, K. A., D. Scharfstein, and J. C. Stein (1993, Dec). Risk management: Coordinating corporate investment and financing policies. *The Journal of Finance 48*(5), 1629–1658.

Gennaioli, N., A. Shleifer, and R. Vishny (2013a). A model of shadow banking. *Journal of Finance 68*(4), 1331–1363.

Gennaioli, N., A. Shleifer, and R. W. Vishny (2013b). Finance and the preservation of wealth. NBER Working Paper.

Gennaioli, N., A. Shleifer, and R. W. Vishny (2014). Money doctors. *Journal of Finance*.

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 (2011). A model of unconventional monetary policy. *Journal of Monetary Economics 58*, 17–34. 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, January). The safe-asset share. NBER WP.

Gorton, G. and A. Metrick (2012). Securitized banking and the run on repo. *Journal of Financial Economics*.

Gorton, G. and G. Pennacchi (1990, March). Financial intermediaries and liquidity creation. *Journal of Finance 45*(1), 49–72.

Gorton, G. and A. Winton (2003). Financial intermediation. In G. M. Constantinides, M. Harris, and R. Stulz (Eds.), *Handbook of the Economics of Finance*, North Holland, pp. 431–552. Elsevier.

Graham, J. R., M. Leary, and M. R. Roberts (2012). A century of capital structure: The leveraging of corporate america. Working Paper Wharton.

Greenwood, J., J. M. Sanchez, and C. Wang (2010, September). Financing development; the role of information costs. *The American Economic Review 100*(4), 1875–1891.

Greenwood, J., J. M. Sanchez, and C. Wang (2012). Quantifying the impact of financial development on economic development. Working Paper, University of Pennsylvani.

Greenwood, R., S. Hanson, and J. Stein (2011). A comparative-advantage approach to government debt maturity. Working Paper Harvard University.

Greenwood, R. and D. Scharfstein (2013). The growth of modern finance. *Journal of Economic Perspectives 27*(2), 3–28.

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 A. T. et al. (Ed.), *The Future of Finance: The LSE Report*. LSE.

Hall, R. E. (2011). The high sensitivity of economic activity to financial frictions. *Economic Journal*.

Hasbrouck, J. (2009, June). Trading costs and returns for u.s. equities: Estimating effective costs from daily data. *Journal of Finance 64*(3), 1445–1477.

He, Z. and A. Krishnamurthy (2012). A model of capital and crises. *Review of Economic Studies*.

Holmström, B. and J. Tirole (1997). Financial intermediation, loanable funds and the real sector. *Quarterly Journal of Economics 112*, 663–691.

Hood, K. (2013, February). Measuring the services of commercial banks in the national income and products accounts. *Survey of Current Business*, 8–19.

Jones, A. and D. Weinberg (2000). The changing shape of the nation's income distribution: 1947-1998. Technical report, US Census Bureau.

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 P. Aghion and S. Durlauf (Eds.), *Handbook of Economic Growth*. Elsevier. Chapter 18.

Kaplan, S. N. and J. Rauh (2010). Wall street and main street: What contributes to the rise in the highest incomes? *Review of Financial Studies 23*(3), 1004–1050.

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.

Kiyotaki, N. and J. Moore (2008). Liquidity, business cycles and monetary policy. mimeo, Princeton.

Krishnamurthy, A. (2009). Amplification mechanisms in liquidity crises. *NBER WP*.

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. Technical report, 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. J. (2000, March). 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.

Midrigan, V. and D. Y. Xu (2011). Finance and misallocation: Evidence from plant-level data. Working Paper, NYU.

Modigliani, F. and M. H. Miller (1958). The cost of capital, corporation finance and the theory of investment. *American Economic Review 48*, 261–297.

Moore, J. (2011). Leverage stacks and the financial system. Ross Prize Lecture.

Moore, K. B. and M. G. Palumbo (2010, June). 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 WP 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. NBER WP 17652.

Philippon, T. (2008). The evolution of the us financial industry from 1860 to 2007: Theory and evidence. NBER WP No. 13405.

Philippon, T. (2012a). Equilibrium financial intermediation. Working Paper NYU.

Philippon, T. (2012b). Finance vs. wal-mart: Why are financial services so expensive? In A. Blinder, A. Lo, and R. Solow (Eds.), *Lessons from the Crisis*.

Philippon, T. and A. Reshef (2012). Wages and human capital in the u.s. financial industry: 1909-2006. *Quarterly Journal of Economics*.

Philippon, T. and A. Reshef (2013). An international look at the growth of modern finance. *Journal of Economic Perspectives 27*(2), 73–96.

Piketty, T. and G. Zucman (2013). Capital is back: Wealth-income ratios in rich countries 1700-2010.

Pozsar, Z., T. Adrian, A. Ashcraft, and H. Boesky (2010, July). Shadow banking. NY Fed Staff Report.

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

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.

Scharfstein, D. (1988). The disciplinary role of takeovers. *The Review of Economic Studies 55*, 185–199.

Scharfstein, D. and A. Sunderam (2011). The economics of housing finance reform. Working Paper Harvard Business School.

Schularick, M. and A. M. Taylor (forthcoming). Credit booms gone bust: Monetary policy, leverage cycles and financial crises, 1870–2008. *American Economic Review*.

Sidrauski, M. (1967). Rational choice and patterns of growth in a monetary economy. *American Economic Review 57*, 534–544.

Sirmans, C. F. and J. D. Benjamin (1990). Pricing fixed rate mortgages: Some empirical evidence. *Journal of Financial Services Research 4*, 191–202. 10.1007/BF00365422.

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. NBER WP 16385.

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 WP No. 14616.