

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

## On the Theory and Measurement of Financial Intermediation

Thomas Philippon\*

March 2012

### Abstract

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

STILL PRELIMINARY

---

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

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

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

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

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

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

The third contribution of the paper is to construct a consistent, theory-based measure of output for the finance

industry. Conceptually, it is useful to distinguish three types of services:<sup>1</sup>

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

Services of type (i) and (ii) involve the creation of various financial assets and liabilities. In the credit market, I measure separately the quantities borrowed by households, farms, non-financial firms, financial firms, and the government. I find that the non-financial corporate credit market is smaller today than it was at its peak of the late 1920s. The most important trends in recent years are the increase in household debt, which exceeds 100% of GDP for the first time in history, and in financial firms' debt, which exceeds that of non-financial firms also for the first time. In the equity market, I measure initial and seasoned offerings. For liquidity I measure deposits, repurchase agreements, and money markets mutual funds. For advising fees I construct a measure of M&A activity. I do not attempt in this paper to measure the informativeness of prices (I discuss the relevant references in Section 5). I then create an output series by aggregating all types of credit, equity issuances, advising and liquidity services produced by the finance industry for the non-financial sector.

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

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

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

---

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

The first strand is the theory of banking and intermediation. While stylized and focused on macroeconomic predictions, the model developed below is consistent with leading theories of financial intermediation, such as Diamond and Dybvig (1983), Diamond (1984), Gorton and Pennacchi (1990), Holmström and Tirole (1997), Diamond and Rajan (2001), and Kashyap, Rajan, and Stein (2002). Gorton and Winton (2003) provide a review of the literature on financial intermediation. The focus of this paper differs from this literature in several ways: (i) the measurement of the costs of intermediation; (ii) the simultaneous modeling of household and corporate finance; and (iii) the use of an equilibrium model to interpret the historical evidence.

There is a large literature on financial development, which I cannot discuss here, except to say that it tends to focus on cross-sectional comparisons of countries at relatively early stages of financial development in order to understand the impact of finance on economic growth (e.g. Rajan and Zingales (1998)) and the determinants of financial development (e.g. La Porta, Lopez-de Silanes, Shleifer, and Vishny (1998), Guiso, Sapienza, and Zingales (2004)). The literature typically focuses on corporate finance (Greenwood, Sanchez, and Wang (2010), Buera, Kaboski, and Shin (2011), Midrigan and Xu (2011)).<sup>2</sup> This paper is more closely related to a recent branch of the literature that seeks to provide risk-adjusted measures of financial productivity (Haldane, Brennan, and Madouros (2010), Basu, Inklaar, and Wang (2011)) and that considers the possibility of inefficient financial development (Glode, Green, and Lowery (2010), Bolton, Santos, and Scheinkman (2011)). Philippon and Reshef (2007) share the historical perspective of this paper, but that paper is purely empirical and focused on the quality and compensation of labor in the finance industry. The large historical changes in the finance share of GDP were first documented and discussed in Philippon (2008), but that paper only focused on corporate credit, which I now estimate to be less than half of the output of the finance industry.<sup>3</sup>

In its account of liquidity services provided by the finance industry, the paper is also related to the classic literature on money and banking. Lucas (2000) provides a benchmark analysis of money demand. A recent branch of this literature has focused on the rise of market-based intermediation, also called shadow banking. Pozsar, Adrian, Ashcraft, and Boesky (2010) document the structure of shadow banking. Gorton and Metrick (2012), Stein (2012), and Gennaioli, Shleifer, and Vishny (2011) emphasize the importance of investors demand for risk free assets. Gorton, Lewellen, and Metrick (2012) argue that much of the shadow banking activity (pooling and tranching) happens to satisfy this demand for risk free assets. I attempt to account for these activities by measuring shadow deposits, such as money market mutual funds and repurchase agreements. The rise of shadow banking also diminishes the relevance of the traditional literature focused on efficiency in banking. That literature did provide

---

<sup>2</sup>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, Demirguc-Kunt, and Levine (2011), the literature uses cross-country data on interest-rate spreads to estimate financing frictions, e.g., Greenwood, Sanchez, and Wang (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).

<sup>3</sup>The paper did not consider household credit, and did not account for liquidity services, which have become important with the rise of the shadow banking system. Another, more technical issue is that the production functions in Philippon (2008) were not appropriate, as discussed in Philippon (2012a).

measures of productivity in banking (see Wang, Basu, and Fernald (2008) for a discussion), but it focused on net interest income, which is only about half of the income of today’s large banks (see the numbers for JP Morgan in the appendix). An important point developed below is that it is difficult to break down the income earned by the finance industry into economically meaningful components.

Finally, it is important to emphasize some important limitations of my analysis. First, it does not deal with financial crises and risk taking. For instance, my output series include all corporate borrowing by Telecom companies in the late 1990s and all subprime and home equity borrowing by households in the mid 2000s. In doing so, I never ask whether borrowing is appropriate or excessive, and I therefore miss the crucial insights of Reinhart and Rogoff (2009). Similarly, I consolidate the earnings of financial intermediaries without controlling for systemic risk taking.<sup>4</sup> Finally, I might be overstating the output of the finance industry because I do not adjust for the role of GSEs in the mortgage market, analyzed by Scharfstein and Sunderam (2011) among others.

The remaining of the paper is organized as follows. In Section 1, I construct my measure of the cost of financial intermediation. Section 2 presents the benchmark model of corporate and household finance to organize the discussion. Section 3 presents measures of output for the finance industry, and computes the unit cost of intermediation. Section 4 presents the quality adjustments. Section 5 discusses the role of information technology, price informativeness, financial derivatives, risk sharing, and trading. Section 6 concludes and discusses avenues for future research.

## 1 Income Share of Finance Industry

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

### 1.1 Raw Data<sup>5</sup>

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

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

---

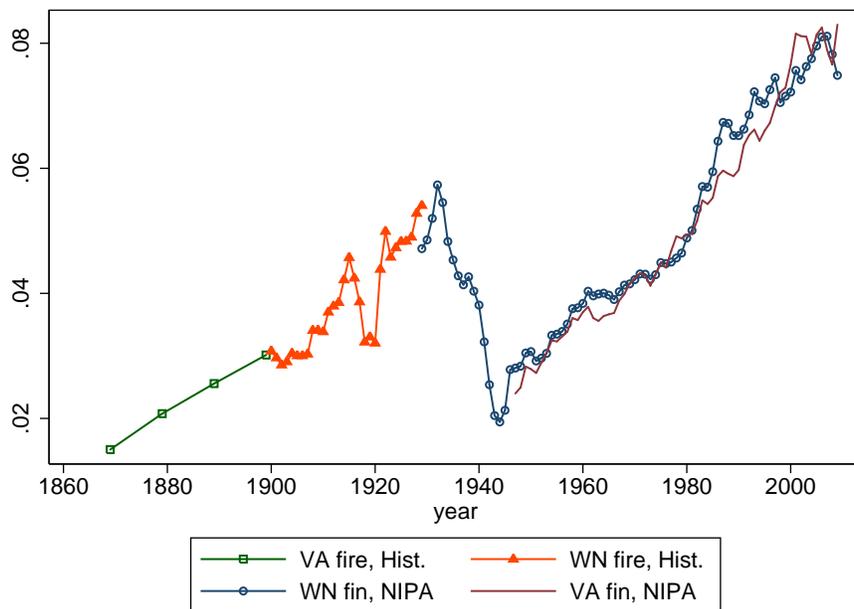
<sup>4</sup>See for instance Adrian and Shin (2008), Krishnamurthy (2009), and Acharya, Pedersen, Philippon, and Richardson (2009) for recent discussions.

<sup>5</sup>The paper uses a lot of data sources. To save space in the paper, all the details regarding the construction of the series are provided in a separate online appendix.

There are two immediate issues:

- Definition of “Finance.” For the most part, financial activities are classified consistently over time (but sub-sectors within finance are not). The main issue is with real estate. The value added of the “real estate” industry includes rents and imputed rents for home owners. Whenever possible, I exclude real estate. In my notations, all variables indexed with “fin” include finance and insurance and exclude real estate. This is not possible before 1929. In this case I use the compensation of employees whenever possible.
- Definition of “Income.” The best measure in theory is total income. In this case,  $\phi$  is the nominal income of the finance industry over the GDP of the US economy. However, this is only acceptable if we can measure the finance industry without imputed rents from the real estate sector. When this is not possible, a good alternative is to use the compensation of employees because the share of real estate is small. In this case,  $\phi$  is the compensation of employees in finance over the total compensation of employees in the US. For the post-war period, the two measures display the same trends, even though annual changes can differ. This simply means that, in the long run, the labor share in the finance industry is the same as the labor share in the rest of the economy. In the short run, of course, profit rates can vary.

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



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

Figure 1 displays various measures of the share of the Finance and Insurance industry in the GDP of the United States estimated from 1870 to 2009. For the period 1947-2009, I use value added and compensation measures

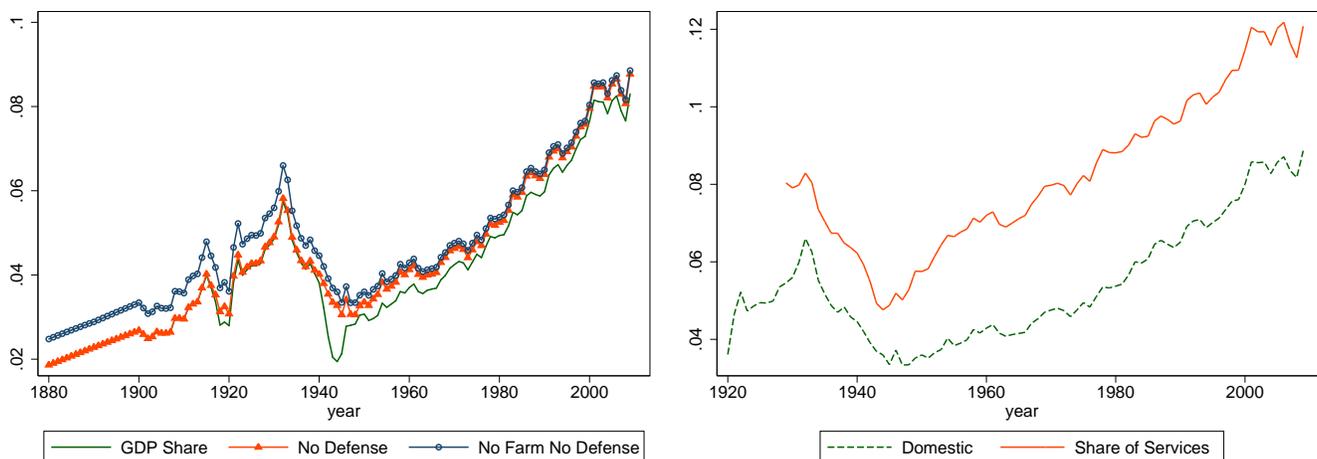
from the Annual Industry Accounts of the United States, published by the Bureau of Economic Analysis (BEA). For 1929-1947, I use the share of employee compensation because value added measures are either unavailable or unreliable. For 1870-1929 I use the Historical Statistics of the United States.<sup>6</sup> More detail regarding the various data sources can be found in Philippon and Reshef (2007) and in the Data Appendix (available online).

The first important point to notice is that the measures are qualitatively and quantitatively consistent. It is thus possible to create one “extended” series simply by appending the older data to the newer ones.<sup>7</sup> The second key point is that finance was smaller in 1980 than in 1925. Given the outstanding real growth over this period, it means that finance size is not simply driven by income per capita.

## 1.2 Adjusted Measures

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

Figure 2: Income Share of Finance (alternative measures)



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

## War and Structural Change

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

<sup>6</sup>Carter, Gartner, Haines, Olmstead, Sutch, and Wright (2006).

<sup>7</sup>Other measures based on Martin (1939) and Kuznets (1941) give also give consistent values.

share of finance in non-farm GDP.

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

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

### **Finance versus Services**

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

### **Globalization and Trade in Financial Services**

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

The right panel of Figure 2 displays the ratio of income minus net exports for finance over non-farm civilian GDP. The figure is almost identical to the previous one. The reason is that the U.S., unlike the U.K. for instance, is not a large exporter of financial services. According to IMF statistics, in 2004, the U.K. financial services trade balance was +\$37.4 billions while the U.S. balance was -\$2.3 billions: the U.S. was actually a net importer. In 2005, the U.K. balance was +\$34.9 billions, and the U.S. balance was +\$1.1 billions.<sup>8</sup>

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

---

<sup>8</sup>There is, of course, some trade within the financial sector, notably between the U.S. and the U.K., but the growth in the GDP share of finance is not due to large net exports.

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

To answer these questions, I introduce financial services to firms and households in general equilibrium model.<sup>9</sup>

## 2 Benchmark Model

I now introduce a model of the economy, which consists of households, a non-financial business sector, and a financial intermediation sector. The economy is non-stationary. The driving force is the labor-augmenting technological progress  $A_t = (1 + \gamma) A_{t-1}$ . The model is a simple particular case of the class of models studied in Philippon (2012a).

I proceed as follows. In the benchmark model borrowers are homogenous. This provides a tight characterization of equilibrium intermediation. I then introduce heterogeneity, which is critical to understand quality adjustments.

### 2.1 Households

The finance industry provides credit as well as liquidity and payment services to households (and savers more generally). In addition, household debt has an important life-cycle component (i.e., mortgages). The model must therefore incorporate these features. To do so, I consider a setup with two types of households: some households are infinitely lived, the others belong to an overlapping generations structure.<sup>10</sup> Households in the model do not lend directly to one another. Savers lend to intermediaries, and intermediaries lend to firms and households.

#### Long-Lived Households

This long-lived household (index 0) owns the capital stock and has no labor endowment. Liquidity services can be modeled using a cash-in-advance framework, or with money in the utility function. I use the later for simplicity, and I specify the utility function as  $u(C_t, M_t) = \frac{(C_t M_t^\nu)^{1-\rho}}{1-\rho}$ , where  $C$  is consumption and  $M$  are holdings of liquid assets. 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

---

<sup>9</sup>The neoclassical growth model can easily be extended to accommodate two sectors and it is well known that the properties of this model depend on the elasticity of substitution between the two sectors (Baumol (1967)). The nominal GDP share of sector  $i$  increases with relative technological progress in sector  $i$  if and only if the elasticity of substitution is less than one. I argue, however, that the traditional multi-sector model is not useful to analyze financial intermediation because it is not the reduced form of any sensible model of financial intermediation (more details can be found in the appendix). This will become evident in the next two sections. The elasticity depends both on the shape of the distribution of borrowers and on the efficiency of the supply of financial services. The standard multi-sector model is not useful to understand the finance industry. Instead, we must explicitly model financial intermediation.

<sup>10</sup>Neither the pure infinite horizon model nor the pure OLG model is adequate. The pure infinite horizon model misses the importance of life-cycle earnings profile. The pure two-periods OLG model is not appealing because households do not actually borrow: the young ones save, and the old ones eat their savings. In addition bequests are of first order empirical importance. The simplest way to capture all these ideas is the mixed model. The interpretation is that the long-lived household has bequest motives, so it is equivalent to an infinitely lived agent. See also Mehra, Piguillem, and Prescott (2011) for a model where household save for retirement over an uncertain lifetime.

$r$  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  $\psi_m$  is the price of liquidity services, and  $S$  are total savings.<sup>11</sup> The Euler equation  $u_C(t) = \beta \mathbb{E}_t [(1 + r_{t+1}) u_C(t+1)]$  can then be written as

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

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

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

## Overlapping Generations

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

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

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

## 2.2 Non Financial Businesses

Non financial output is produced with constant returns technology  $Y_t = F(A_t n_t, K_t)$ . The capital stock  $K_t$  depreciates at rate  $\delta$ . Firms start each period with capital in place  $x K_t$  and obtain the remaining  $(1 - x) K_t$  from intermediaries (or from issuances underwritten by intermediaries). Let  $\psi_{k,t}$  be the price of corporate financial intermediation. Non financial firms therefore solve the following program  $\max_{n,K} F(A_t n, K) - (r_t + \delta + (1 - x) \psi_{k,t}) K - W_t n$ . Capital demand equates the marginal product of capital to its user cost:

$$\frac{\partial F}{\partial K}(A_t n_t, K_t) = r_t + \delta + (1 - x) \psi_{k,t}. \quad (1)$$

---

<sup>11</sup>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-Douglas 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.

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

$$A_t \frac{\partial F}{\partial n} (A_t n_t, K_t) = W_t. \quad (2)$$

Finally, I assume that the production function is Cobb-Douglas.

**Assumption:**  $F(A_t n_t, K_t) = (A_t n_t)^\alpha K_t^{1-\alpha}$

&& Philippon (2012a) discusses the details. CD is a good assumption. Qualitatively different results only happen when elasticity of substitution between capital and labor is above 6. &&

### 2.3 Intermediation Equilibrium

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

Financial services are produced with capital and labor with constant returns to scale. Philippon (2012a) discusses in details the implications of various production functions, when financial intermediaries explicitly hire capital and labor. These issues are not central here, and I therefore assume financial services are produced from final goods with a constant marginal cost  $\zeta$ . The quantity of financial services is given by

$$Y^\phi = \mu_c B_t^c + \mu_m M_t + \mu_k B_t^k, \quad (3)$$

where  $B_t^k = (1-x)K_t$  and  $\mu_i$  is the intermediation intensity of market  $i$ . For instance, if it is more complicated to monitor entrepreneurs than households, we would have  $\mu_k > \mu_c$ . The parameters  $\mu$ 's are convenient to describe various prediction of the model. For instance, a decrease in  $\mu_k$  can be interpreted either as an improvement in the creditworthiness of businesses, or as an improvement specific to corporate finance (i.e., that leaves the marginal cost  $\zeta$  unchanged, and does not directly affect household finance). It is immediate from (3) that for each market  $j = c, m, k$  we have

$$\psi_j = \zeta \mu_j. \quad (4)$$

An *equilibrium* in this economy is a sequence of the various prices and quantities listed above such that households choose optimal levels of credit and liquidity; financial and non financial firms maximize profits; the labor market clears

$$n_t = 1;$$

the capital market clears

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

Aggregate GDP in this economy is defined by  $\bar{Y}_t \equiv Y_t + \zeta Y_t^\phi$  and the finance share of GDP discussed in Section 1 is

$$\phi_t \equiv \frac{\zeta Y_t^\phi}{Y_t + \zeta Y_t^\phi}$$

**Proposition 1.** *Under HI, there is a unique balanced growth path where the finance share of GDP  $\phi$ , the unit cost of financial intermediation  $p^\phi$ , and the financial ratios  $M/Y$ ,  $B/Y$  and  $K/Y$  are constant. The equilibrium has the following features*

(i) *Improvements in corporate finance increase output, the real wage, and the capital-output ratio, household debt increases proportionally to GDP, and liquidity increases more than proportionally;*

(ii) *Improvements in household finance increase household debt, consumption and liquidity, but do not affect the real wage;*

(iii) *Improvements in liquidity management increase consumption, but do not affect household debt or the real wage;*

(iv) *Demand shifts  $(x, \eta)$  increase the finance income share while supply shifts  $(\zeta)$  have an ambiguous impact on the income share;*

*Proof.* See Appendix. □

That homogeneity 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 a detailed discussion). Regarding liquidity demand, balanced growth comes from the assumed preferences, as discussed in Lucas (2000).

From now on, I use lower-case letters for their de-trended variables, i.e. variables scaled by the current level of technology. For instance, for capital I write  $k \equiv \frac{K_t}{A_t}$ , for consumption of agent  $i$   $c_i \equiv \frac{C_{i,t}}{A_t}$ , and for the the productivity adjusted wage:  $w \equiv W_t/A_t$ . Let us discuss the main features of the equilibrium. On the BGP,  $M$  grows at the same rate as  $C$ . The Euler 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

$$\beta(1+r) = (1+\gamma)^\theta. \tag{5}$$

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

With Cobb-Douglass technology, the capital labor ratio is

$$\frac{k}{n} = \left( \frac{1 - \alpha}{r + \delta + (1 - x)\psi_k} \right)^{\frac{1}{\alpha}}.$$

Since  $n = 1$  in equilibrium, this is also the aggregate stock of capital. Non financial GDP is  $y = k^{1-\alpha}$ . The real

wage is

$$w = \alpha \left( \frac{k}{n} \right)^{1-\alpha} = \alpha y.$$

The wage is increasing in the efficiency of corporate finance. Note that the capital stock and corporate intermediation are independent of consumer credit and liquidity. The fact that consumer credit does not crowd out corporate credit is not a general property, it comes from the simplifying assumption that finance uses no labor. In the general case, there is crowding out (see Philippon (2012a)).

Given the interest rate in (5), the Euler equation of short lived households becomes simply

$$c_1 = (1 - \psi_c)^{\frac{1}{\theta}} c_2. \quad (6)$$

In addition, we have  $\psi_m m = \nu c$  for each cohort. The budget constraints are therefore  $(1 + \nu) c_1 = \eta_1 w + (1 - \psi_c) b$  and  $(1 + \nu) c_2 = \eta_2 w - \frac{1+r}{1+\gamma} b$ . We can then use the Euler equations and budget constraints to compute the borrowing by young households:

$$b_c = \frac{(1 - \psi_c)^{\frac{1}{\theta}} \eta_2 - \eta_1}{1 - \psi_c + (1 - \psi_c)^{\frac{1}{\theta}} \frac{1+r}{1+\gamma}} w. \quad (7)$$

If  $\psi_c$  is 0, we have  $c_1 = c_2$ , and consumption grows at the rate  $\gamma$  for all agents in the economy. From the perspective of current consumption, borrowing costs act as a tax on future labor income. If  $\psi_c$  is too high, no borrowing takes place and the consumer credit market collapses. The bigger the ratio  $\eta_2/\eta_1$  the larger the borrowing. For instance, increased years schooling generates more borrowing, and a larger financial sector. Improvements in corporate finance increase household debt  $b_c$  because they increase  $w$ , but  $b_c/y$  remain constant since  $w = \alpha y$ .

The “real” output of the finance industry is

$$y^\phi = \mu_c b_c + \mu_m m + \mu_k b_k$$

We already know  $b_c$  and  $b_k$ . Finally, liquidity demand is

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

We already know  $\psi_m = \zeta \mu_m$ , so we only need to measure aggregate consumption as

$$c = \frac{1}{1 + \nu} (w - \psi_c b + (r - \gamma) k)$$

Improvements in corporate finance increase liquidity demand because they increase the consumption output ratio.

When  $\psi_k$  goes down,  $k/y$  goes up while  $b/y$  is unchanged, therefore  $\nu c_0/y$  goes up.

The “nominal” output of the finance industry is  $\zeta y^\phi = \psi_c b + \psi_m m + \psi_k b_k$  and the finance share of GDP is

$$\phi = \frac{\zeta y^\phi}{y + \zeta y^\phi}$$

One important point is that, under HI, the model does not predict an income effect, i.e., just because a country becomes richer does not mean that it should spend a higher fraction of its income on financial services. I will test directly HI in Section 3.

### 3 Output and Unit Cost

In this section I construct empirical proxies for  $\frac{m}{y}$ ,  $\frac{b_c}{y}$ ,  $\frac{b_k}{y}$  as well as other elements of what the finance industry does.

#### 3.1 Credit Markets

Figure 3 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.<sup>12</sup>

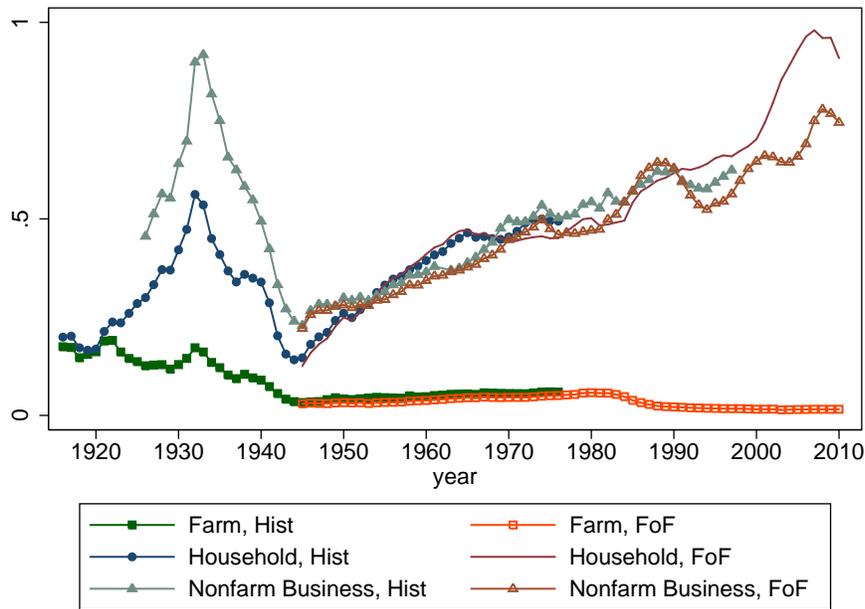
In the theory outlined earlier, there is no distinction between outstanding assets and new issuances. In the data the two can be different and it is useful to consider stocks and flows separately. Figure 4 shows the issuances of corporate bonds by non-financial corporations. Note that issuances collapse 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 4 also shows a measure of household debt issuance.

To extend the credit series before 1920, I use the balance sheets of financial firms as well as data on home mortgages provided by Schularick and Taylor (forthcoming). I measure assets on the balance sheets of commercial banks, mutual banks, savings and loans, federal reserve banks, brokers, and life insurance companies. I define total assets as the sum of assets of all these financial firms over GDP. I use this series to extend the total non-financial debt series (households & non corporates, farms, corporates, government). I regress total credit on total assets and use the predicted value to extend the credit series.

---

<sup>12</sup>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 3: 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.

### 3.2 Equity Market

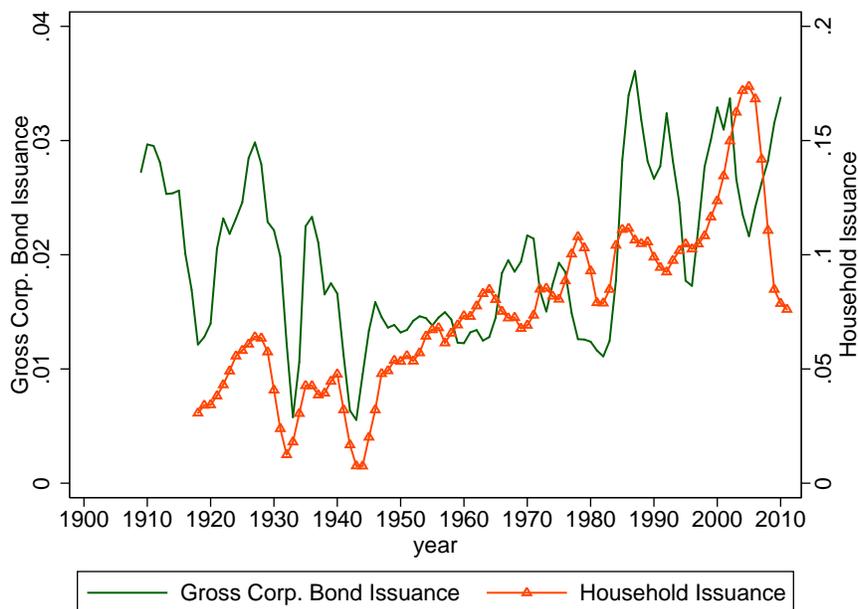
The equity market is difficult to deal with because of valuation effects. Stocks, unlike bonds, are recorded at market value. The ratio of the market value of equity to GDP can fluctuate without any intermediation services (i.e., without any issuance of equity). Another problem is that net issuances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

### 3.3 Money and Liquidity

In addition to credit (on the asset side of banks), households, firms and local governments benefit from payment and liquidity services (on the liability side of banks and money market funds). For households, I use the total currency and deposits, including money market fund shares, held by households and nonprofit organizations. The left panel of Figure 6 shows the evolution of this variable.

An important issue in the measurement of liquidity provision is the rise of the shadow banking system. Gorton, Lewellen, and Metrick (2012) argue that a significant share recent activities in the financial sector was aimed at creating risk free assets with money-like features. For firms (incorporated or not), I follow Gorton, Lewellen, and

Figure 4: 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

Metrick (2012) and I treat repos as shadow deposits. The series is thus the sum of checkable deposits and currency, time and savings deposits, money markets mutual funds shares, and repos (by non financial firms).<sup>13</sup>

### 3.4 Aggregation

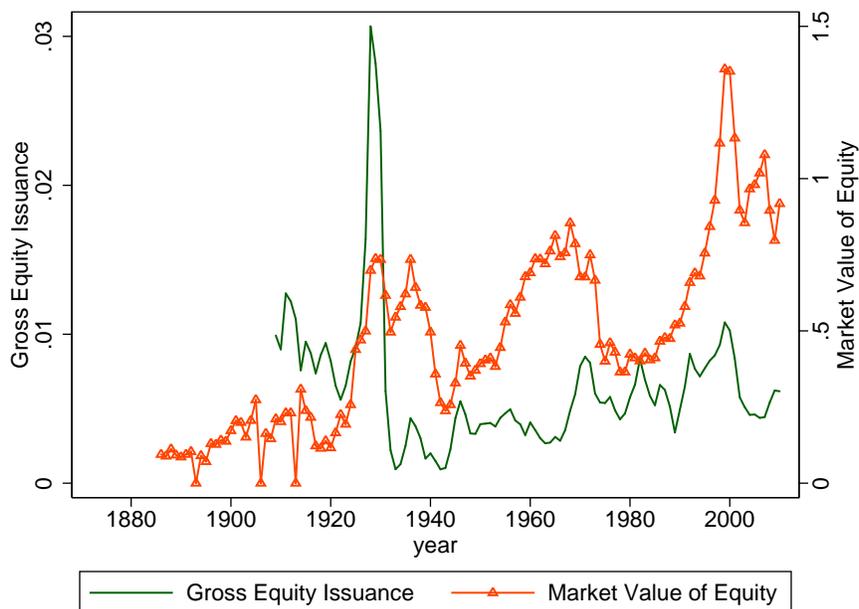
The model presented in Section 2 suggests to measure the output of the finance industry as the weighted sum of assets intermediated. I normalize the weight on consumer credit to one. Finance output is then  $y^\phi = \mu_c b_c + \mu_m m + \mu_k b_k + \mu_g b_g$  where  $b_k$  is corporate intermediation,  $b_c$  is household borrowing,  $m$  are deposits, and  $b_g$  is government debt. An important issue is that I cannot measure the GDP share linked to these assets separately, so I cannot directly estimate the parameters  $\mu$ 's.<sup>14</sup> I will therefore first assume that the  $\mu$ 's are constant, while allowing the average  $\zeta$  to move over time. I will later provide quality adjustment that change the relative  $\mu$ 's.

Households save  $S$  at a gross return of  $1 + r$ , while liquid assets yield  $(1 + r) / (1 + \psi_m)$ . Conceptually,  $\psi_m$  is the cost of creating liquid assets and also the price investors are willing to pay for liquidity. It could be measured as a redemption fee for investors. These fees are around 2%. They are consistent with the trading costs incurred by mutual funds upon withdrawals (see Chen, Goldstein, and Jiang (2010) for a discussion). It can also be measured as an opportunity cost to savers. Over the period 1990-2008, Table 1 shows that the opportunity cost of cash was

<sup>13</sup>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.

<sup>14</sup>This is not a simple accounting issue. Financial tasks are deeply intertwined. Insurance companies and pension funds perform their own independent credit analysis. Banks act as market makers. Investment banks behave as hedge funds. In addition, the mapping from industry to tasks has changed over time. XX

Figure 5: 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)

Table 1: U.S. Interest Rates and Returns, 2002-2011

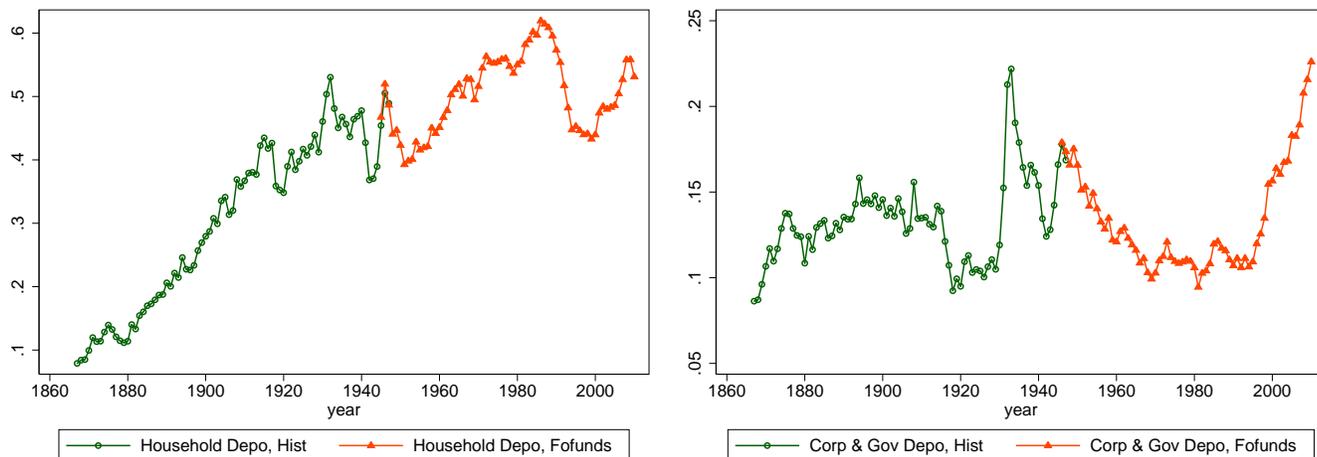
	Value (%)
Average 3M Treasury Bills	1.79
Average 1M Certificate of Deposits	2.14
Average 1Y Gov. Bond	2.10
Average 10Y Gov. Bond	3.95
Average Aaa Corporate	5.47
Average Prime Bank Loan	5.02
Average 30Y Conventional Mortgage	5.71
Average Baa Corporate	6.64
Vanguard Prime Money Market Fund Return	1.82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

more than 4%. Cash is of course a relatively small fraction of liquid assets. Over the past 10 years, Vanguard Prime Money Market has returned 1.82% with an expense ratio of 0.20, while Vanguard Short Term Treasury has returned 3.65% (with an expense ratio of 0.22).<sup>15</sup> The spread between these two essentially risk free instrument is 1.8%. I will therefore assume that  $\psi_m = 2\%$ .

Next I need to compare the intermediation requirements of households and businesses. Corporate intermediation itself is made of equity and debt financing, hence  $\mu_k b_k = \mu_k^d d_k + \mu_k^e e_k$ , where  $e_k$  is corporate equity and  $d_k$  is corporate debt. For corporate issuances, Altinkilic and Hansen (2000) report fees of 3% to 4% for equity and about 1% for bonds. For flows, I therefore assume  $\mu_k^e = 3.5\mu_k^d$ . 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,

<sup>15</sup>Vanguard data accessed on March 11, 2012.

Figure 6: Deposits and Shadow Deposits



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

and around 2.46% for equity (see Appendix). Of course, these are for large borrowers. For households we can look at the mortgage market. Sirmans and Benjamin (1990) report fees of 0.5% to 0.7%. For other types of consumer credit these fees are certainly larger. Table 1 suggests that once issued, high quality corporate and consumer debt trades at similar prices. As a benchmark, I will assume that  $\mu_k^d = \mu_c$  so that it is equally difficult to extend credit to firms or to households.

I now need to construct two series for each type of output, one of flows of new assets, and one of levels of outstanding assets. Before doing so, I need to discuss M&A activities.

## M&As

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

## Level Measure

The data suggests will also assume  $\mu_k^d = \mu_c$  and  $\psi_m = 2\%$ . It will turn out that the average cost of intermediation is also around 2%. Anticipating this, and for ease of exposition, I therefore simply assume that  $\mu_m = \mu_c$ , and this will be consistent with the evidence. There is also the issue of the debt of the government. On the one hand, it is risk-free and liquid, and it might actually help the functioning of financial markets (Krishnamurthy and Vissing-

Jorgensen (2010), Greenwood, Hanson, and Stein (2011)). So one option would be to ignore government debt, i.e., to assume that it does not require intermediation services. On the other hand, there is some duration risk, and it needs to be traded. Somewhat arbitrarily, I assume that government debt requirements are 1/10 of that of private debt:  $\frac{\mu_g}{\mu_c} = 0.1$ . Finally, for the level measure, I treat equity and debt similarly. The level measure is therefore:

$$y_{level}^{\phi} = b_c^{level} + d_k^{level} + e_k^{level} + m^{level} + 0.1b_g^{level} + y_{M\&A}^{level}$$

## Flow Measure

Flows correspond to issuances. I calibrate the relative weights using underwriter fees as explained earlier. When I do not have a separate measure of flows, I assume a runoff rate of 10% and I create the flow measure from the level series.

$$y_{flow}^{\phi} = b_c^{flow} + d_k^{flow} + 3.5e_k^{flow} + m^{flow} + 0.1b_g^{flow} + y_{M\&A}^{flow}$$

In practice, I have data on  $d_k^{flow}$  only for corporate bond issuances. I assume that similar quantities of loans are issued, so that  $d_{k,corp}^{flow} = 2d_{k,corp,bonds}^{flow}$ , which is in line with recent years. Similarly, I miss the the non-corporate sector, so I use  $d_k^{flow} + 3.5e_k^{flow} = 1.5 * (d_{k,corp}^{flow} + 3.5e_{k,corp}^{flow})$ .

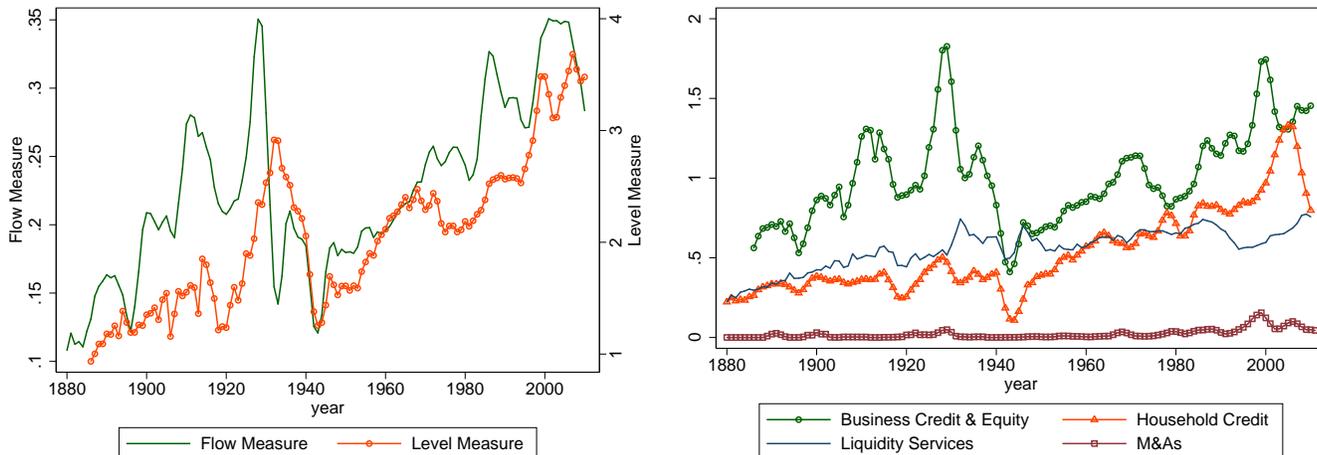
## Composite measures

As explained above, I construct two output series for the finance industry. One using the flows (gross issuances over GDP) and one using the levels (debt and equity over GDP). Note that both are relevant in theory. Screening models apply to the flow of new issuances, while monitoring models apply to the stocks. Trading applies to both. The two series are displayed in the left panel of Figure 7. The production of financial services increases steadily until WWI, and rapidly after 1919 until 1929. It collapses during the great depression and WWII. It increases steadily until 1975 and more randomly afterwards. The flow and level measures share the same trends, but there are clear differences at medium frequencies. The flow measure is more stationary than the level measure. The flow measure collapses quickly during the great depression while the level measure peaks later in 1932-33 (this is exacerbated by deflation). A similar pattern seems to emerge during and after the great recession of 2008-2009.

I now wish to combine the flow and level measures. Since there is no reason to prefer one to the other, it seems logical to put equal weights on the two measures. On average, in the post-1950 period (where the data is most reliable), the ratio of the flow measure to the level measure is 0.062. In other words, the level measure is about 16 times the flow one. I therefore set  $\lambda = \bar{y}_{flow}^{\phi} / \bar{y}_{level}^{\phi} = 1/10$  and I construct the composite measure as

$$y^{\phi} = \frac{1}{2} \left( y_{level}^{\phi} + \frac{y_{flow}^{\phi}}{\lambda} \right) \quad (8)$$

Figure 7: Output of Finance Industry



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

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

The right panel of Figure 7 presents the composite measures corresponding to 4 broad functions discussed earlier: credit and equity intermediation services to firms (farms, corporate, non-corporate), credit intermediation services to households, liquidity services to both, and M&A activities. In each case, the composite measure is based on the stocks and flow measures aggregated as in (8). Note that the liquidity and M&A measures are not in fact composite measures since I do not have independent flow measures for deposits or level measures for M&As. It is clear from this figure that credit intermediation for firms and households are the most volatile series. There is also a significant increase in liquidity services in the 2000s. M&As play a significant role mostly in the 1990s.

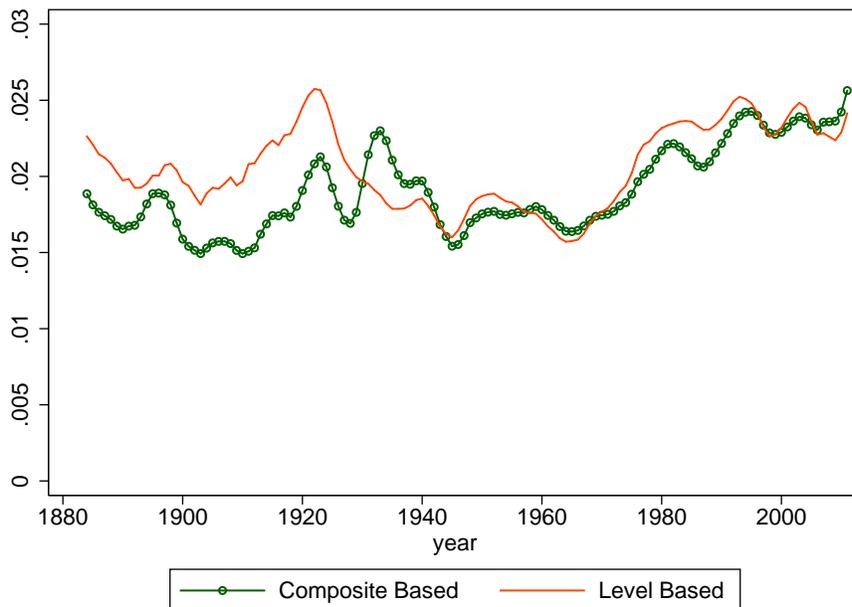
### 3.5 Unadjusted Unit Cost Measures

Figure 8 estimates the cost of financial intermediation, defined as income divided by output. For output, I use the composite measure of equation (8) as well as the level measures. Figure 8 is the main contribution of the paper. It brings together the theory of Section 2 and the historical/empirical work of Section 3. There are two main points to take away. The first is that the ratio is relatively stable. Recall that we started from a series in Figure 1 that fluctuates by a factor of 5 (9% relative to less than 2%). All the debt, deposit and equity series also vary a lot over time. But their ratio, properly scaled, seems quite stable. On Figure 8 it stays roughly between 1.5% and 2.5% over 130 years. One must also keep in mind that the model ignores business cycles and assumes constant real interest rates.

The second main point is that the finance cost index has been trending upward, especially since the 1970s. This

is counter-intuitive. If anything, the technological development of the past 40 years (IT in particular) should have disproportionately increased efficiency in the finance industry. How is it possible for today's finance industry not to be significantly more efficient than the finance industry of John Pierpont Morgan? I conclude that Figure 8 presents a puzzle. In the next section, I show how quality adjustments go some way towards solving this puzzle.

Figure 8: Unadjusted Intermediation Cost Index



Notes: total intermediation costs divided by production of credit assets, equity and deposits. Composite measure uses average of levels and flows.

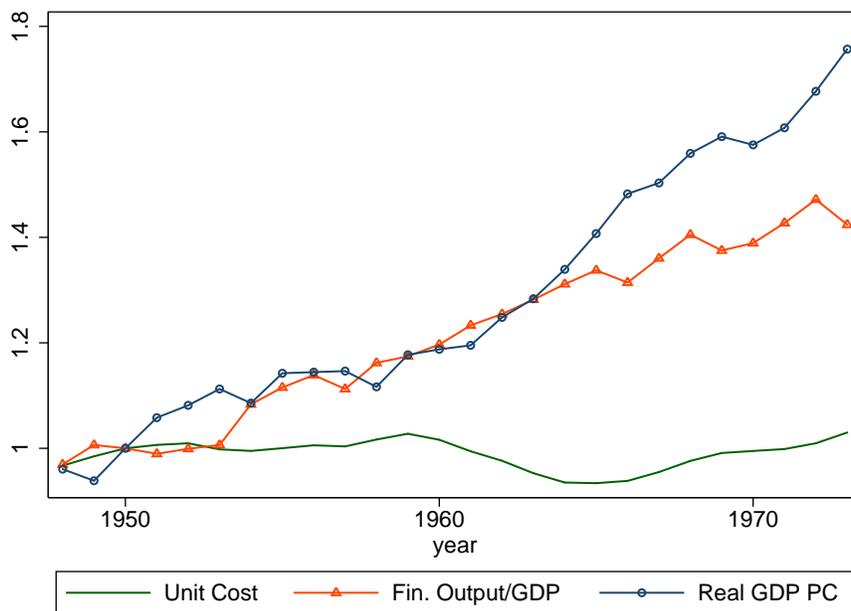
### 3.6 Evidence of Constant Returns to Scale

As shown earlier, under HI, the theory does not predict that the finance share of GDP should mechanically increase with income. The right panel of figure 9 presents evidence consistent with HI. It uses the period 1947-1973, for two reasons. First, the post-war data is reliable. Second, as I will discuss shortly, quality adjustments are not important over this period. Since these adjustments are difficult to implement, it is more convincing to present the evidence without them.

From 1950 to 1980, real GDP per-capita increases by 80% and the ratio of finance output to GDP increase by more than 50%, but my estimate of the unit cost of intermediation remains roughly constant (all series are presented as ratios of their 1950 values). In other words, people are a lot richer, financial markets are a lot larger, but the unit cost does not change. This provides fairly strong support for HI.<sup>16</sup>

<sup>16</sup>Bickenbach, Bode, Dohse, Hanley, and Schweickert (2009) show that the income share of finance has remained remarkably constant in Germany over the past 30 years. More precisely, using KLEMS for Europe (see O'Mahony and Timmer (2009)) one can see that the finance share in Germany was 4.3% in 1980, 4.68% in 1990, 4.19% in 2000, and 4.47% in 2006.

Figure 9: Constant Returns to Scale



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

## 4 Heterogeneity and Quality Adjustments

### 4.1 Theory

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

Let us therefore consider a model with heterogeneity and decreasing returns at the firm level.<sup>17</sup> Let  $k$  be the (endogenous) number of firms, and let  $n = \bar{n}/k$  be employment per-firm. Each firm operates  $A_t$  units of capital and hires  $n$  workers to produce  $A_t f(n)$  units of output, where  $f$  is increasing and concave. Decreasing returns come from the fact that capital is fixed at the firm level. Firms choose employment to maximize (detrended) net income

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

<sup>17</sup>Decreasing returns in production are required to make room for heterogeneity since with constant returns borrowers that have even a slight financial disadvantage would not be able to enter.

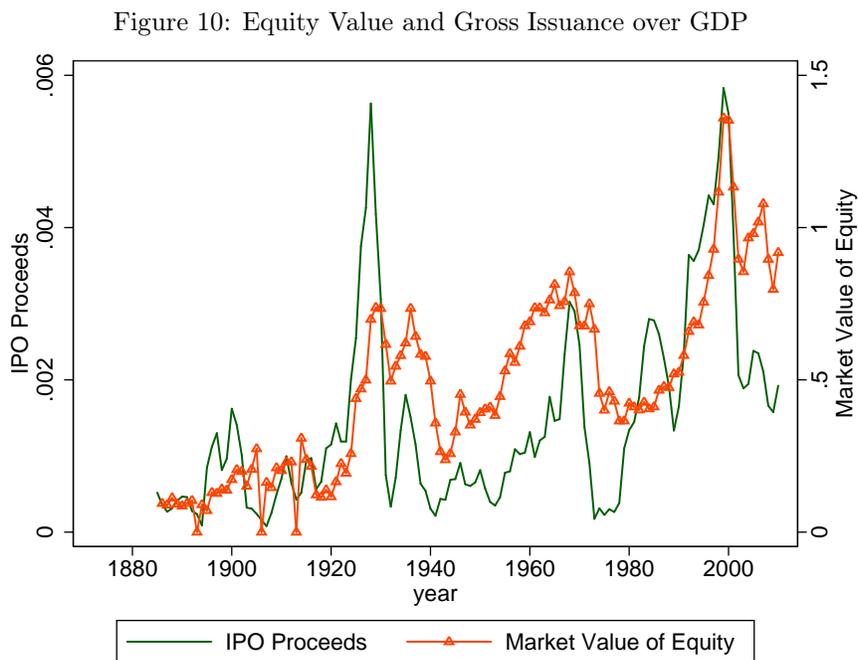
Each firm hires  $n$  employees such that  $f'(n) = w$ . The labor demand schedule is the decreasing function  $n(w)$ .

Macroeconomic adjustment to the stock of capital takes place at the extensive margin, i.e., by firms' entry (and exit) decisions. Firms and households differ in their financing needs. [TBC]

## 4.2 Quality Adjustments

The theory presented above makes it clear that the quantities of intermediation should be adjusted for the difficulty of monitoring/screening borrowers. To be concrete, suppose that borrowers become harder to screen. Then a given amount of lending requires more intermediation. As a result, the income share of finance will increase. Without quality adjustment, my measure would register a spurious increase in intermediation cost (i.e., a spurious decrease in financial efficiency). I consider two separate adjustments, one for corporate borrowing, and one for household borrowing. The two adjustments are displayed on Figure 11.

Figure 10 shows the IPO and market value measures. As argued by Jovanovic and Rousseau (2001), the IPO market of the 1920s was remarkably active, even compared to the one of the 1990s: the IPO firms are of similar ages, and the proceeds are comparable. By contrast, the market value of GDP has an upward trend, in part due to increases in price earnings ratios, and in part due to an increase in the stock market listing.



Notes: Market Value of non-financial corporate firms from the Flow of Funds and from CRSP. IPO is a three-year centered moving average of IPO proceeds over GDP. Sources are Jovanovic and Rousseau (2001) and Ritter (2011).

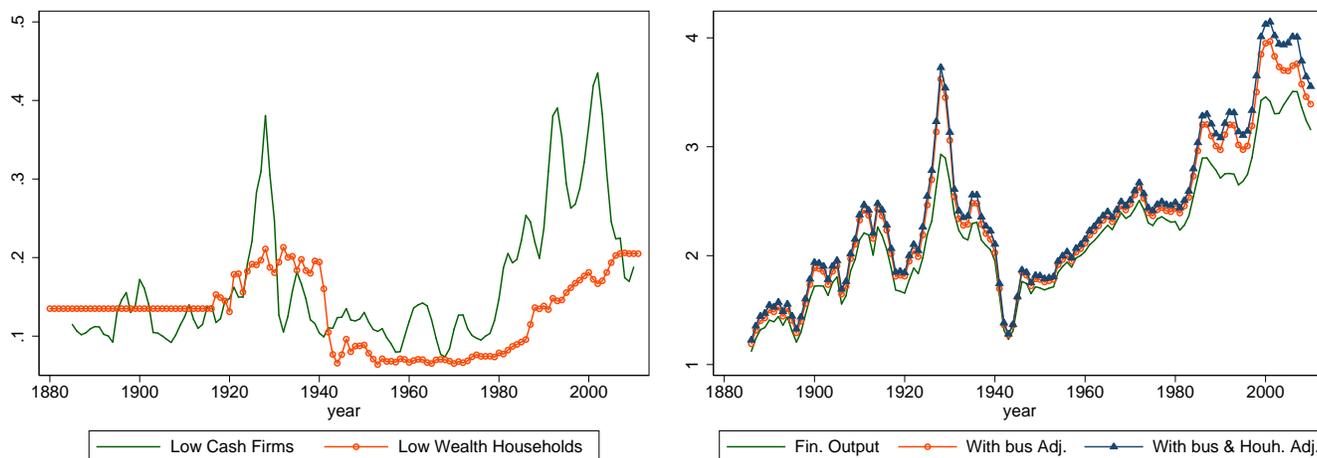
For corporate borrowing, the appendix provides a simple adjustment based on the aggregate importance of cash-poor firms. Philippon (2008) uses Compustat to construct an empirical proxy for this share, namely the share of aggregate investment that is done by firm who must borrow more than 3/4 of their capital spending. This

measure is thus available from 1950 to 2010. To extend the series in the past, I use IPO proceeds. In the post-war period, when both series are available, they are highly correlated. Once this is done, I use a calibrated version of the corporate finance model to compute the required adjustment. The model suggests an adjustment of 1.25 times the share of low cash firms, with a normalization so that aggregate monitoring equals the composite measure of firm intermediation when the share is zero. If the share of low cash firms is 20%, the model says that the true production of intermediation services is 25% higher than the composite credit measure. The adjustment is therefore

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

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

Figure 11: Quality Adjustments



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

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

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

To extend the series I assume that the share of low wealth household is proportional to inequality, and I use data

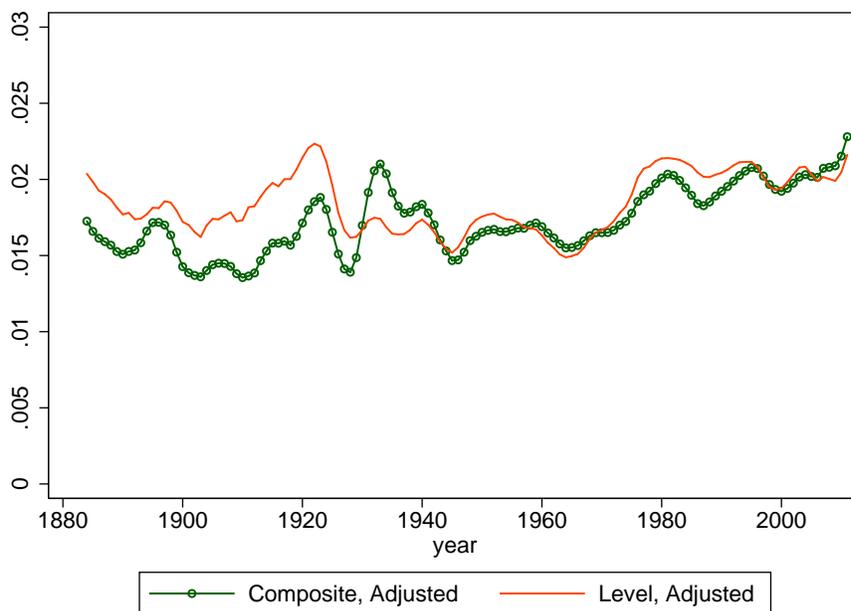
from Piketty and Saez (2003) to extend the series.

### 4.3 Unit Cost

Figure 12 estimates the cost of financial intermediation, defined as income divided by output. For output, I use the composite measure of equation (8). Figure 12 is the main contribution of the paper. It brings together the theory and the historical/empirical work of Section 3. There are two main points to take away. The first is that the ratio is remarkably stable. Recall that we started from a series in Figure 1 that fluctuates by a factor of 5 (9% relative to less than 2%). All the debt, deposit and equity series also vary a lot over time. But their ratio, properly scaled, seems quite stable. On Figure 12 it stays roughly between 1.5% and 2.5% over 130 years. One must also keep in mind that the model ignores business cycles and assumes constant real interest rates.

The second main point is that the finance cost index has been trending upward, especially since the late 1970s. This is counter-intuitive. If anything, the technological development of the past 40 years (IT in particular) should have disproportionately increased efficiency in the finance industry. How is it possible for today's finance industry not to be significantly more efficient than the finance industry of John Pierpont Morgan? I conclude that Figure 12 presents a puzzle for future research. In the next (and last) section, I discuss some recent or ongoing research that might shed light on this puzzle.

Figure 12: Quality-Adjusted Cost Index



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

## 5 Discussion

### 5.1 Information Technology

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

### 5.2 Price Informativeness and Risk Sharing

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

#### **Informativeness of Prices**

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

#### **Risk Management and Derivatives**

Another benefit of financial intermediation is risk sharing. Risk sharing can affect firms and households. At the firm level, risk sharing is commonly called risk management. Better risk management would, in equilibrium, mostly

translate into lower cost of fund, more issuances and more investment. This first effect would be captured by our measures of debt and equity issuances. Better risk management could also increase TFP if high productivity projects are also riskier.

The market for financial derivatives is extremely large. Since these contracts are in zero net supply, however, they do not enter directly into the calculation of output for the finance industry. How should we account for these contracts? One thing is clear: it would not make sense to count derivatives at face value.<sup>18</sup> In terms of economic theory, derivatives can add real value in one of two ways: (i) risk sharing; (ii) price discovery. Risk sharing among intermediaries would not create a bias in my measurements, however. It would simply lead to lower borrowing costs and cheaper financing.<sup>19</sup> Therefore, the only bias from derivative contracts must come from better risk sharing or price discovery among non-financial borrowers.

The correct way to measure the value added of derivatives is to measure directly the informativeness of prices, or the welfare gains from risk sharing among non-financial firms and households. I have already argued that there is no evidence of improvement in informativeness of prices. I am also not aware of any direct evidence suggesting significant improvements in corporate risk management. One obvious index, that of precautionary savings by businesses, suggest even the opposite: corporate cash holdings have increased over the past 30 years. There is also no direct evidence of credit derivatives leading to better risk management, and it is commonly believed that hedging represents only a small fraction of all trades in the CDS market.<sup>20</sup>

### **Risk Sharing & Consumption Smoothing**

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

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

---

<sup>18</sup>For instance, consider the following example. Without derivatives, corporation A borrows from bank B and bank B retains the credit and duration risks on its books. With derivatives, bank B buys insurance against credit risk from fund C using a CDS. The sum of B and C holds exactly the same risk. Absent other frictions, the two models are exactly equivalent.

<sup>19</sup>To see why let us go back to the simple example. Suppose there are frictions that rationalize why B and C should be separate entities, and why they gain from trading with each other (i.e., B has a comparative advantage at managing duration risk, and C at managing credit risk). Then the existence of CDS contracts can improve risk sharing among intermediaries, lower the risk premia, and lead to a decrease in the borrowing costs of A. Hence, with free entry, the total income going to intermediaries B+C would decrease. This could then increase the demand for borrowing, as explained 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.

<sup>20</sup>Also remember that the liquidity measure captures deposits and repos contracts by non financial firms. So cash management is accounted for by my measures.

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

### 5.3 Trading

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

One proximate cause might be trading. Trading costs have decreased (Hasbrouck (2009)), but trading volumes have increased even more. In addition, the costs of active fund management are large. French (2008) estimates that investors spend 0.67% of asset value trying (in vain, by definition) to beat the market. French's calculations are only for the equity market. In Figure 12, a drop in the intermediation cost index of 50 to 60 basis points would indeed bring it back towards its historical average. With output at 4 times GDP, this suggests that about 2% of GDP, or about \$280 billions annually, are either wasted or at least difficult to account for.

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

## 6 Concluding Remarks

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

# Appendices

## A Proof of Proposition 1

I use the convention of upper-case letters for variables with trends, and lower-case letters for their de-trended counterparts. For instance, for capital I write  $k_t \equiv \frac{K_t}{A_t}$  and for consumption  $c_t \equiv \frac{C_t}{A_t}$ .

Long-lived households own the capital stock, or in detrended form

$$s = b + (1 + \gamma) \bar{k}.$$

The interest rate is pinned down by (10). The intermediation production function defines a constant marginal cost  $\chi^\phi$  which must equal the price of financial services:  $p^\phi = \chi^\phi(w, r + \delta + \psi_k)$ . Using (4) for business capital, we therefore have

$$\psi_k = \mu_k \chi^\phi(w, r + \delta + \psi_k)$$

which leads to an intermediation cost of capital  $\psi_k$  that increasing in the real wage  $w$  and decreasing in the efficiency of intermediation. Similarly in the non-financial business sector we have

$$\chi(w, r + \delta + \psi_k) = 1$$

This pins down the wage  $w$ .

We can then detrend the market clearing condition as

$$y^\phi = F^\phi(n^\phi, k^\phi) = \mu_c b + \mu_m m + \mu_k \bar{k} \quad (9)$$

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

$$\beta(1 + r) = (1 + \gamma)^\theta. \quad (10)$$

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

The Euler equation for short-lived households written with detrended consumption is

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

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

$$c_1 = (1 - \psi_c)^{\frac{1}{\theta}} c_2. \quad (11)$$

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

We can use the Euler equations and budget constraints to compute the borrowing by young households:

$$b = \frac{(1 - \psi_c)^{\frac{1}{\theta}} \eta_2 - \eta_1}{1 - \psi_c + (1 - \psi_c)^{\frac{1}{\theta}} \frac{1+r}{1+\gamma}} w. \quad (12)$$

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

Given constant returns, we can write output as  $Y_t = A_t F(n, k_t)$ . I focus on the balanced growth path with constant growth rate  $\gamma$ .

where  $w$  is the productivity adjusted wage:  $w \equiv W_t/A_t$

To close the model, we obtain the consumption of the long lived savers ( $c_0$ ) from the resource constraint. Since  $\psi_m m_i = \nu c_i$  for all agents, we have  $y = (1 + \nu)(c_0 + c_1 + c_2) + \psi_c b + (\gamma + \delta + \psi_k)(k + k^\phi)$ . Using the budget constraints of the young and old agents, we get

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

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

## B Monitoring and Corporate Finance

For an analysis of incentive provision and monitoring via the takeover market, see Scharfstein (1988), Bolton and Scharfstein (1990).

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

	Table 2: Events within Time $t$			
	(i) Investment	(ii) Effort	(iii) Labor Market	(iv) Payoffs
Household (mass 1)	Capital $K_t$	.	Earn $W_t$	Earn $(1 + r) K$
Entrepreneur (mass $k$ )	Project size $A_t$	Effort+Monitoring	Hire $n$	$\gamma A$
Key Equation	$K_t = k A_t$	(IC) constraint	$A_t \pi(w)$	Value: $(\pi + 1 - \delta) A_t$

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

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

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

### B.1 Credit Constraints

A project is successful only if the entrepreneur behaves well. If not, the project fails, investors recover nothing, while the entrepreneur enjoys the private benefit  $\gamma A$ . One interpretation is that the entrepreneur consumes a fraction  $\eta$  of the capital stock, while the remaining  $1 - \eta$  is wasted. Investors must make sure that the entrepreneur behaves well. Let  $sA$  be the payoff to the entrepreneur in case of success. The incentive compatibility constraint for the entrepreneur is therefore  $sA > \eta A$ , or simply  $s > \eta$ . The pledgeable income of the project is the maximum value that the lenders can receive  $(\pi - \eta) A + (1 - \delta) A$ . The project can be financed if and only if  $\pi - \eta > r + \delta$ . In the first best economy under constant returns to scale we have of course  $\pi = r + \delta$ . Therefore there will always be rationing relative to first best. Therefore the wage is pinned down by

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

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

## B.2 Monitoring

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

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

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

$$\mu\psi < \gamma$$

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

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

It is clear that finance is constant on the Balanced Growth Path.<sup>21</sup>

## C Accounting

Long-lived households own the capital stock

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

Adding up the budget constraints we have

$$W_t + (1+r)S_{t-1} + (1-\psi_c)B_t - (1+r)B_{t-1} - \psi_m M_t = C_{0t} + C_{1t} + C_{2t} + S_t$$

The two sides of GDP are

$$\begin{aligned}Y_t &= W_t + (r + \delta + \psi_k)K_t \\ Y_t &= K_{t+1} + C_{0t} + C_{1t} + C_{2t} - (1 - \delta - \psi_k)K_t + \psi_m M_t + \psi_c B_t\end{aligned}$$

Combining them we get

$$K_{t+1} + C_{0t} + C_{1t} + C_{2t} = W_t + (1+r)K_t - \psi_c B_t - \psi_m M_t$$

Combining with the budget constraint and capital market equilibrium we get

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

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

---

<sup>21</sup>It is also straightforward to introduce moral hazard on the side of banks. In this case, banks must have incentives to monitor and they must receive a minimum amount  $y^\mu$  in the good state. This is the model of Holmström and Tirole (1997).

On the balanced growth path we have

$$\begin{aligned} y &= (1 + \gamma)k + c_0 + c_1 + c_2 - (1 - \delta - \psi_k)k + \psi_m m + \psi_c b \\ y &= (\gamma + \delta + \psi_k)k + c_0 + c_1 + c_2 + \psi_m m + \psi_c b \end{aligned}$$

We can write

$$(1 + \nu)(c_0 + c_1 + c_2) = w + (r - \gamma)k - \psi_c b$$

The budget constraint of short lived households is

$$(1 + \nu)(c_1 + c_2) = w + (1 - \psi_c)b - \frac{1 + r}{1 + \gamma}b = w - \psi_c b - \frac{r - \gamma}{1 + \gamma}b$$

so

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

## D Complete Model with Heterogeneity

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

### D.1 Firms

With Cobb-Douglas technology, we get net income

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

Monitoring demand

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

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

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

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

$$kn(w) = 1$$

Let  $k_h$  be the (exogenous) number of high cash firms (these are infra-marginal so they earn pure profits). Then we get

$$k_l = k - k_h$$

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

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

and corporate borrowing

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

Output is simply

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

And corporate finance income is

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

Using Cobb-Douglass, we have

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

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

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

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

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

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

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

## D.2 Households

Long-term households pin down the interest rate

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

Money demand is

$$\psi_m m = \nu c$$

Short-lived households consumption equation is such that

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

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

Using  $w = \alpha y$ :

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

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

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

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

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

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

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

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

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

and therefore we get

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

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

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

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

### Household adjustment [to be completed]

I introduce inequality among households. Half of the households draw a relatively high income  $\eta_1^h$  and the other half a low income  $\eta_1^l$ .

Poor households need to borrow more.

## E JP Morgan 2010

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

## 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.
- ADRIAN, T., AND H. S. SHIN (2008): "Financial Intermediary Leverage and Value at Risk," Federal Reserve Bank of New York Staff Reports, 338.
- AGUIAR, M., AND M. BILS (2011): "Has Consumption Inequality Mirrored Income Inequality?," Working Paper, University of Rochester.
- ALTINKILIC, O., AND R. S. HANSEN (2000): "Are There Economies of Scale in Underwriting Fees? Evidence of Rising External Financing Costs," *Review of Financial Studies*, 13, 191–218.
- BAI, J., T. PHILIPPON, AND A. SAVOV (2011): "Have Financial Prices Become More Informative?," .
- BAKER, M., AND J. WURGLER (2000): "The Equity Share in New Issues and Aggregate Stock Returns," *Journal of Finance*.
- BASU, S., R. INKLAAR, AND J. C. WANG (2011): "The Value of Risk: Measuring the Services of U.S. Commercial Banks," *Economic Inquiry*, 49(1), 226–245.
- BAUMOL, W. J. (1967): "Macroeconomics of Unbalanced Growth: The Anatomy of the Urban Crisis," *American Economic Review*, 57, 415–426.
- BECK, T., A. DEMIRGUC-KUNT, AND R. LEVINE (2011): *Financial Structure and Economic Growth: A Cross-Country Comparison of Banks Markets, and Development*chap. The Financial Structure Database, pp. 17–80. MIT Press, Cambridge.
- BEKAERT, G., C. R. HARVEY, AND R. LUMSDAINE (2002): "Dating the Integration of World Capital Markets," *Journal of Financial Economics*, 65, 203–249.
- BICKENBACH, F., E. BODE, D. DOHSE, A. HANLEY, AND R. SCHWEICKERT (2009): "Adjustment After the Crisis: Will the Financial Sector Shrink?," Kiel Policy Brief.
- BOLTON, P., T. SANTOS, AND J. SCHEINKMAN (2011): "Cream Skimming in Financial Markets," Working Paper, Columbia University.
- BOLTON, P., AND D. SCHARFSTEIN (1990): "A Theory of Predation Based on Agency Problems in Financial Contracting," *The American Economic Review*, 80(1), 93–106.
- 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 Millennium 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.
- DIAMOND, D. W. (1984): "Financial Intermediation and Delegated Monitoring," *Review of Economic Studies*, 51, 393–414.
- DIAMOND, D. W., AND P. H. DYBVIK (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.
- ESTEVADEORDAL, A., B. FRANTZ, AND A. M. TAYLOR (2003): "The Rise and Fall of World Trade, 1870-1939," *The Quarterly Journal of Economics*, 118(2), 359–407.
- FRENCH, K. R. (2008): "Presidential Address: The Cost of Active Investing," *The Journal of Finance*, 63(4), 1537–1573.
- GENNAIOLI, N., A. SHLEIFER, AND R. VISHNY (2011): "A Model of Shadow Banking," Working Paper, Harvard University.
- GERARDI, K. S., H. ROSEN, AND P. WILLEN (2010): "The Impact of Deregulation and Financial Innovation on Consumers: The Case of the Mortgage Market," *Journal of Finance*, 65(1), 333–360.

- GERTLER, M., AND P. KARADI (2009): “A Model of Unconventional Monetary Policy,” Working Paper NYU.
- GERTLER, M., AND N. KIYOTAKI (2010): “Financial Intermediation and Credit Policy in Business Cycle Analysis,” Working Paper, NYU.
- GLODE, V., R. C. GREEN, AND R. LOWERY (2010): “Financial Expertise as an Arms Race,” forthcoming in *Journal of Finance*.
- GORTON, G., S. LEWELLEN, AND A. METRICK (2012): “The Safe-Asset Share,” NBER Working Paper.
- GORTON, G., AND G. PENNACCHI (1990): “Financial Intermediaries and Liquidity Creation,” *Journal of Finance*, 45(1), 49–72.
- GORTON, G., AND A. WINTON (2003): “Financial Intermediation,” in *Handbook of the Economics of Finance*, ed. by G. M. Constantinides, M. Harris, and R. Stulz, pp. 431–552, North Holland. Elsevier.
- GORTON, G. B., AND A. METRICK (2012): “Securitized Banking and the Run on Repo,” *Journal of Financial Economics*.
- GREENWOOD, J., J. M. SANCHEZ, AND C. WANG (2010): “Financing Development; The Role of Information Costs,” *The American Economic Review*, 100(4), 1875–1891.
- (2012): “Quantifying the Impact of Financial Development on Economic Development,” Working Paper, University of Pennsylvania.
- GREENWOOD, R., S. HANSON, AND J. STEIN (2011): “A Comparative-Advantage Approach to Government Debt Maturity,” Working Paper Harvard University.
- GUIO, L., P. SAPIENZA, AND L. ZINGALES (2004): “The Role of Social Capital in Financial Development,” *American Economic Review*, 94, 526–556.
- HADAS, E. (2011): “Commodity prices are failing New Zealand test,” *Financial Times*.
- HALDANE, A., S. BRENNAN, AND V. MADOUROS (2010): “What is the contribution of the financial sector: Miracle or mirage?,” in *The Future of Finance: The LSE Report*, ed. by A. T. et al. LSE.
- HASBROUCK, J. (2009): “Trading Costs and Returns for U.S. Equities: Estimating Effective Costs from Daily Data,” *Journal of Finance*, 64(3), 1445–1477.
- HOLMSTRÖM, B. (1979): “Moral Hazard and Observability,” *Bell Journal of Economics*, 10(1), 74–91.
- HOLMSTRÖM, B., AND J. TIROLE (1997): “Financial Intermediation, Loanable Funds and the Real Sector,” *Quarterly Journal of Economics*, 112, 663–691.
- JOVANOVIĆ, B., AND P. L. ROUSSEAU (2001): “Why Wait? A Century of Life Before IPO,” *AER paper and proceedings*, 91, 336–341.
- JOVANOVIĆ, B., AND P. L. ROUSSEAU (2005): “General Purpose Technologies,” in *Handbook of Economic Growth*, ed. by P. Aghion, and S. Durlauf. Elsevier, Chapter 18.
- KASHYAP, A., R. RAJAN, AND J. STEIN (2002): “Banks as Liquidity Providers: An Explanation for the Coexistence of Lending and Deposit-Taking,” *Journal of Finance*, 57, 33–73.
- KRISHNAMURTHY, A. (2009): “Amplification Mechanisms in Liquidity Crises,” *NBER Working Paper*.
- KRISHNAMURTHY, A., AND A. VISSING-JORGENSEN (2010): “The Aggregate Demand for Treasury Debt,” Working Paper Kellogg School of Management.
- KUZNETS, S. (1941): “National Income and Its Composition, 1919-1938,” Discussion paper, National Bureau of Economic Research.
- LA PORTA, R., F. LOPEZ-DE SILANES, A. SHLEIFER, AND R. W. VISHNY (1998): “Law and Finance,” *Journal of Political Economy*, 106, 1113–1155.
- LUCAS, R. E. J. (2000): “Inflation and Welfare,” *Econometrica*, 68(2), 247–274.
- LUCAS, R. E. J., AND N. L. STOKEY (1987): “Money and Interest in a Cash-in-Advance Economy,” *Econometrica*, 55(3), 491–513.
- MARTIN, R. F. (1939): *National Income in the United States, 1799-1938*. National Industrial Conference Board.
- MAYER, C., AND K. PENCE (2008): “Subprime Mortgages: What, Where, and to Whom?,” Staff Paper Federal Reserve Board.

- MEHRA, R., F. FIGUILLEM, 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.
- MOORE, K. B., AND M. G. PALUMBO (2010): "The Finances of American Households in the Past Three Recessions: Evidence from the Survey of Consumer Finances," Staff Paper Federal Reserve Board.
- OBSTFELD, M., AND A. M. TAYLOR (2002): "Globalization and Capital Markets," NBER Working Paper 8846.
- ODEAN, T. (1998): "Volume, Volatility, Price, and Profit when all traders are above average," *Journal of Finance*, 53, 1887–1934.
- O'MAHONY, M., AND M. P. TIMMER (2009): "Output, Input and Productivity Measures at the Industry Level: The EU KLEMS Database," *The Economic Journal*, 119(538), F374–F403.
- PAGNOTTA, E., AND T. PHILIPPON (2011): "Competing on Speed," Working Paper NYU.
- PHILIPPON, T. (2008): "The Evolution of the US Financial Industry from 1860 to 2007: Theory and Evidence," NBER Working Paper No. 13405.
- (2012a): "Equilibrium Financial Intermediation," Working Paper NYU.
- (2012b): "Finance vs. Wal-Mart: Why are Financial Services so Expensive?," in *Lessons from the Crisis*, ed. by A. Blinder, A. Lo, and R. Solow.
- PHILIPPON, T., AND A. RESHEF (2007): "Wages and Human Capital in the U.S. Financial Industry: 1909–2006," NBER Working Paper 13437.
- PIKETTY, T., AND E. SAEZ (2003): "Income Inequality in the United States, 1913–1998," *Quarterly Journal of Economics*, 118(1), 1–39.
- POZSAR, Z., T. ADRIAN, A. ASHCRAFT, AND H. BOESKY (2010): "Shadow Banking," NY Fed Staff Report.
- RAJAN, R. G., AND L. ZINGALES (1998): "Financial Dependence and Growth," *American Economic Review*, 88, 559–586.
- REINHART, C. M., AND K. S. ROGOFF (2009): *This Time Is Different: Eight Centuries of Financial Folly*. Princeton University Press.
- RHODES-KROPP, 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," .
- WANG, J. C., S. BASU, AND J. G. FERNALD (2008): "A General-Equilibrium Asset-Pricing Approach to the Measurement of Nominal and Real Bank Output," NBER Working Paper No. 14616.