Has the US Finance Industry Become Less Efficient?
On the Theory and Measurement of Financial Intermediation

By Thomas Philippon

A quantitative investigation of financial intermediation in the United States over the past 130 years yields the following results: (i) the finance industry’s share of gross domestic product (GDP) is high in the 1920s, low in the 1960s, and high again after 1980; (ii) most of these variations can be explained by corresponding changes in the quantity of intermediated assets (equity, household and corporate debt, liquidity); (iii) intermediation has constant returns to scale and an annual cost of 1.5–2 percent of intermediated assets; (iv) secular changes in the characteristics of firms and households are quantitatively important. (JEL D24, E44, G21, G32, N22)

This paper is concerned with the theory and measurement of financial intermediation. The role of the finance industry is to produce, trade, and settle financial contracts that can be used to pool funds, share risks, transfer resources, produce information, and provide incentives. Financial intermediaries are compensated for providing these services. The income received by these intermediaries measures the aggregate cost of financial intermediation. This income is the sum of all spreads and fees paid by nonfinancial agents to financial intermediaries and it is also the sum of all profits and wages in the finance industry. This cost of financial intermediation affects the user cost of external finance for firms who issue debt and equity, and the costs for households who borrow or use asset management services.

* Stern School of Business, New York University, 44 West Fourth Street, New York, NY 10012, NBER, and CEPR (e-mail: tphilipp@stern.nyu.edu). This has been a very long project. The first draft dates back to 2007, with a focus on corporate finance, but without the long-term historical evidence. This paper really owes a lot to other people, academics and non-academics alike. Darrell Duffie, Robert Lucas, Raghuram Rajan, Jose Scheinkman, Robert Shiller, Andrei Shleifer, and Richard Sylla have provided invaluable feedback at various stages of this project. Boyan Jovanovic, Peter Rousseau, Moritz Schularick, and Alan Taylor have shared their data and their insights, and I have greatly benefited from discussions with Lewis Alexander, Patrick Bolton, Markus Brunnermeier, John Cochrane, Kent Daniel, Douglas Diamond, John Geanakoplos, Gary Gorton, Robin Greenwood, Steve Kaplan, Anil Kashyap, Ashley Lester, Andrew Lo, Guido Lorenzoni, Andrew Metrick, William Nordhaus, Matthew Rhodes-Kropf, David Robinson, Kenneth Rogoff, Tano Santos, David Scharstein, Hyun Shin, Jeremy Stein, Gillian Tett, Wallace Turbeville, and Luigi Zingales, as well as seminar participants at Stanford, Yale, NYU, Harvard, Chicago, Princeton, and the Paris School of Economics. I also thank Paul Krugman for his discussion at the 2011 NY Area Monetary conference, Axelle Ferriere, Peter Gross, Andrea Prestipino, Markus Sihvonen, Robert Turley, and Shaojun Zhang for research assistance, and the Smith Richardson Foundation for its financial support. The author declares that he has no relevant or material financial interests that relate to the research described in this paper. Two online appendices describe calibrations and data sources.

† Go to http://dx.doi.org/10.1257/aer.20120578 to visit the article page for additional materials and author disclosure statement.
In equilibrium, the user cost of external finance is the sum of the rate of returns to saver \((r)\) and the unit cost of financial intermediation \((\psi)\):

\begin{equation}
\text{user cost of finance} = r + \psi.
\end{equation}

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

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

A second motivation is to shed light on the transformation of the finance industry that has occurred since the 1970s. For instance, we would certainly like to know if the move away from traditional banking and toward an “originate-and-distribute” model has lowered the cost of funds for households and businesses. This is precisely what \(\psi\) should measure. Similarly, if derivatives markets lower hedging costs, their growth should translate into lower funding costs and higher asset values. Any debate about financial regulation is also a debate about \(\psi\), since it involves a trade-off between safety and efficiency.\(^2\) The broader point here is that learning about \(\psi\) is important from a positive perspective and from a normative perspective.

This paper seeks to define and measure financial intermediation. It treats the finance industry as a black box and attempts to measure what goes in, what comes out, and how much the whole system costs. It is important, however, to understand just how difficult the measurement problem is. A simple illustration is given in Figure 1. At the prevailing market rates of 5 and 7 percent, borrowers (firms or households) want to borrow $100, and savers want to save $100. To flow back and forth between savers and borrowers, the funds go through financial intermediaries. These intermediaries need $2 to pay their wage bill and rent the necessary capital. In the terminology of this paper, the quantity of intermediated assets is $100, the intermediation cost is $2, and therefore \(\psi = 2\) percent. Figure 1 presents two fundamentally equivalent ways to organize financial intermediation. In traditional banking, intermediation occurs under one roof: the bank makes a loan, keeps it on its books, and earns a net interest income. This income compensates for the cost of screening and monitoring the borrower and for managing the duration and credit

---

\(^1\) See Greenwood, Sanchez, and Wang (2010); Buera, Kaboski, and Shin (2011); and Midrigan and Xu (2014) for recent analyses of financial development and growth. In addition, much of recent work has focused on the macroeconomic consequences of a sudden increase in \(\psi\), and on the link between \(\psi\) and intermediary capital, leverage, and liquidity. Cúrdia and Woodford (2010); Gertler and Kiyotaki (2010); Hall (2011); Christiano and Ikeda (2011); and Corsetti et al. (2011) study the impact of negative shocks to financial intermediation, building on the classic contribution of Bernanke, Gertler, and Gilchrist (1999). Gertler and Karadi (2011); He and Krishnamurthy (2012); and Moore (2011) focus on liquidity. This paper only deals with the long-term evolution of \(\psi\), but the value of \(\psi\) in normal times is an important parameter even if one is interested in the deviations from its long-term trend.

\(^2\) Take the debate about capital adequacy ratios for banks for instance. Bank leverage has risen substantially since the late nineteenth century, as discussed in Haldane, Brennan, and Madouros (2010) among others. If Modigliani and Miller’s (1958) proposition holds, as Admati et al. (2011) argue, we should not expect a link between \(\psi\) and bank leverage, but if the proposition fails we might expect a downward drift in \(\psi\) over time.
risk of the loan. In the originate and distribute model, by contrast, there is a daisy
chain of intermediation. Many transactions occur inside the black box, with total
face values potentially much larger than $100. There is no simple measure of net
interest income as in the traditional model: there are origination fees, asset manage-
ment fees, trading profits, etc. But the sum of wages and profits for all intermediaries
is still $2, and the quantity of intermediated assets seen from outside the black box
is still $100.

Three steps are required, then, to understand financial intermediation, and these
steps determine the structure of the paper: (i) measure the income of financial
intermediaries; (ii) define and construct the quantity of intermediated assets; and
(iii) compute the unit cost of intermediation and perform quality adjustments. For
the purposes of this paper, the real difficulty lies in the heterogeneity among intermediated assets. Savings vehicles are heterogeneous: perhaps households save $50 in liquid claims with a return of 4 percent and $50 in illiquid claims with a return of 6 percent. The average (expected and risk adjusted) return is 5 percent, but intermediaries must bear the cost of creating and maintaining liquid claims. Borrowers are also heterogeneous: young firms and blue-chip companies, wealthy households and poor households. Changes in the composition of borrowers affect the cost of intermediation, while improvements in financial intermediation give access to credit to borrowers who were previously priced out.

The first contribution of this paper is empirical. Figure 2 shows that the quantity of financial intermediation varies dramatically over time. The first series, constructed in Section I, is the income of financial intermediaries divided by GDP. The income share grows from 2 to 6 percent from 1880 to 1930. It shrinks to less than 4 percent in 1950, grows slowly to 5 percent in 1980, and then increases rapidly after 1980.

Given these large historical variations in the finance income share, it is natural to ask if there are commensurate changes in the quantity of intermediated assets. Section II relies on a simple extension of the neoclassical growth model

---

3 Note that I use the recently updated data from the Bureau of Economic Analysis. The 2014 Comprehensive Revision of Industry Accounts has led to downward revisions in the estimated value added of finance and insurance for the 1997–2012 period. See http://www.bea.gov/industry/index.htm. This revision has occurred between the first version of this paper and the current version, so readers of earlier versions might notice some differences.

4 We can think of the finance industry as providing three types of services: (i) liquidity (means of payments, cash management); (ii) transfer of funds (pooling funds from savers, screening, and monitoring borrowers); (iii) information (price signals, advising on M&As). Financial firms typically produce a bundle of such services. For instance, risk management uses all three types of activities. Services of type (i) and (ii) typically involve the creation of various financial assets and liabilities. This classification is motivated by the mapping between theory
as an accounting framework for household finance, corporate finance, and liquidity provision. The size of the various markets varies significantly over time. The most important trend in credit markets in recent years is the increase in household debt. The business credit market is relatively large in the 1920s, small in the 1960s, and large again after 1980, although not as large as in the late 1920s. I also measure the market value of outstanding equity and the flows of initial and seasoned offerings. Deposits, repurchase agreements, and money markets mutual funds are used to measure liquidity services. After aggregating the various types of credit, equity issuances, and liquid assets into one measure, I obtain the quantity of financial assets intermediated by the financial sector displayed in Figure 2.

I can then divide the income of the finance industry by the quantity of intermediated assets to obtain a measure of the unit cost $\psi$. Figure 3 shows that this unit cost is around 1.5–2 percent and relatively stable over time. In other words, I estimate that it costs $0.02 per year to create and maintain $1 of intermediated financial asset. I also find clear evidence that financial services are produced under constant returns to scale. For instance, from 1947 to 1973 (a period of stable growth without major financial crises), real income per capita increases by 80 and real financial assets by 250 percent, but my estimate of the unit cost of intermediation remains remarkably constant.

The raw measure of Figure 3, however, does not take into account changes in the characteristics of borrowers. The final contribution of the paper is to perform quality adjustments to the quantity of intermediated assets. The 1920s and 1990s

and measurement discussed throughout the paper. It differs a little bit from that of Merton (1995). I do not attempt in this paper to measure the informativeness of prices. This issue is tackled by Bai, Philippon, and Savov (2011). See the discussion at the end of Section III.
are times of entry by young and risky firms, and Jovanovic and Rousseau (2005) have shown that this pattern is related to waves of technological innovation. In the household credit market, relatively poor households have gained access to credit in recent years. In both cases, the challenge is to account for the fact that these borrowers require more intermediation per unit of credit extended. I rely on theory to make the required quality adjustments, which appear to be quantitatively important. According to my calculations, in the 1990s, the raw measure of intermediation underestimates the true quantity by about 25 percent. Given the size of intermediated markets, the failure to adjust for quality would represent a measurement error of the order of one GDP. Figure 3 shows that the adjusted unit cost is more stable than the unadjusted one.

Even with the quality adjustment, however, I find that the unit cost of intermediation is about as high today as it was at the turn of the twentieth century. Improvements in information technologies do not appear to have led to a significant decrease in the unit cost of intermediation. Explaining this puzzle is an active area of research, some of which is discussed at the end of Section III.

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

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

There is a large literature on financial development, which I do not have room to discuss here, except to say that it tends to focus on cross-sectional comparisons of countries at relatively early stages of financial development in order to understand the impact of finance on economic growth (e.g., Rajan and Zingales 1998), and the determinants of financial development itself (e.g., La Porta et al. 1998; Guiso, Sapienza, and Zingales 2004). The literature typically focuses on corporate finance (Greenwood, Sanchez, and Wang 2010; Buera, Kaboski, and Shin 2011; Midrigan and Xu 2014); except Mehra, Piguillem, and Prescott (2011), who study intermediation in a model where households save for retirement over an uncertain lifetime.\(^5\)

\(^5\)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 US finance industry. As a result, both theory and measurement must be expanded. For instance, following Beck, Demirgüç-Kunt, and Levine (2001), the literature uses cross-country data on interest-rate spreads to estimate financing frictions (e.g., Greenwood, Sanchez, and Wang 2013). To study the US finance industry, it is important to recognize that non-interest income (fees, trading revenues, etc.) is now the
This paper is more closely related to a recent branch of the literature that seeks to provide risk-adjusted measures of financial productivity (Wang, Basu, and Fernald 2009; Haldane, Brennan, and Madouros 2010; Basu, Inklaar, and Wang 2011).

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

Finally, there is an emerging literature on the growth of the finance industry. Philippon and Reshef (2012) share the historical perspective of this paper but focus on the composition of the finance labor force. Greenwood and Scharfstein (2013) provide an illuminating study of the growth of modern finance in the United States. They show that two activities account for most of this growth over the past 30 years: asset management and the provision of household credit. For asset management, they uncover an important stylized fact: individual fees have typically declined but the allocation of assets has shifted toward high fee managers in such a way that the average fee per dollar of assets under management has remained roughly constant. While most of the existing work has focused on the United States, Philippon and Reshef (2013) and Bazot (2013) provide evidence for other countries.

A second set of papers offers theoretical explanations for the growth of finance documented in this paper and in the empirical papers discussed above. There are two main stylized facts to explain: the size of finance (see Figure 2) and the unit cost (see Figure 3). Regarding this second stylized fact, a puzzle seems to be that the unit cost has not declined despite obvious improvements in information technologies. As a result, the income received by financial intermediaries might be unexpectedly high. In Glode, Green, and Lowery (2012), an “arms race” can occur as agents try to protect themselves from opportunistic behavior by (over-)investing in financial expertise. In Bolton, Santos, and Scheinkman (2011), cream skimming in one market lowers assets quality in the other market and allows financial firms to extract excessive rents. In Pagnotta and Philippon (2011) there can be excessive investment in trading speed because speed allows trading venues to differentiate and charge higher prices. Gennaioli, Shleifer, and Vishny (2014b) propose an alternative interpretation for the relatively high cost of financial intermediation. In their model, trusted intermediaries increase the risk tolerance of investors, allowing them to earn higher returns. Because trust is a scarce resource, improvements in information technology do not necessarily lead to a lower unit cost.

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 banks’ creation of safe assets is driven by investors’ liquidity demand (all these points are discussed in details below).

6 The large historical changes in the finance share of GDP were first documented and discussed in Philippon (2008), but that paper only focused on corporate credit. The paper did not consider household credit, and did not account for liquidity services, which have become important with the rise of the shadow banking system.
The other fact to explain is the size of the finance industry. Since the unit cost appears to be roughly constant, the question becomes: how do we explain the large historical variations in the ratio of intermediated assets over GDP? This paper documents that the income share of the finance industry is roughly equal to 2 percent of the ratio of intermediated assets over GDP, but it does not seek to explain the size of intermediated assets. Gennaioli, Shleifer, and Vishny (2014a) propose an explanation. They argue that the growth of finance can be explained by the rise of the wealth to income ratio, documented in Piketty and Zucman (2014) for several countries. The driving force is a slowdown in aggregate growth which leads, along the transition path, to an increase in the capital output ratio. If the unit cost of intermediation does not fall as the capital output ratio increases, then the income share of the finance industry increases.

Let me end this introduction with an important caveat: this paper does not analyze financial crises. The model assumes that credit markets clear via prices, not via covenants or quantity restrictions as we often see during crises. In the model, borrowers can be priced out, but inefficient rationing does not occur. Similarly, the model does not study whether borrowing is appropriate or excessive, whether financial intermediaries take on too much aggregate risk, nor whether government interventions create moral hazard.

The remainder of the paper is organized as follows. Section I estimates the income of financial intermediaries. Section II computes the quantity of intermediated assets. Section III implements quality adjustments and discusses the role of information technology and price informativeness. Section IV concludes.

I. Income Share of Finance

In this section, I present the first main empirical fact: the evolution of the total cost of financial intermediation in the United States over the past 140 years. As argued in the introduction, there is no simple way to break down the income earned by the finance industry into economically meaningful components. For instance, insurance companies and pension funds perform credit analysis, fixed income trading provides liquidity to credit markets, and securitization severs the links between assets held and assets originated. From a historical perspective, these issues are compounded by regulatory changes in the range of activities that certain intermediaries can provide. Rather than imposing arbitrary interpretations on the data, I therefore focus on a consolidated measure of income earned by all financial intermediaries, irrespective of whether they are classified as private equity funds, commercial banks, insurance companies, or anything else.

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

8 See for instance Adrian and Shin (2014); Reinhart and Rogoff (2009); Krishnamurthy (2010); Acharya et al. (2009); and Scharfstein and Sunderam (2011) for recent discussions of these issues.
A. Raw Data

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

\[
\frac{y_t^f}{y_t} = \frac{\text{Value Added of Finance Industry}}{\text{GDP}}.
\]

Conceptually, the best measure is value added, which is the sum of profits and wages. Whenever possible, I therefore use the GDP share of the finance industry, i.e., the nominal value added of the finance industry divided by the nominal GDP of the US economy. One issue, however, is that before 1945 profits are not always properly measured and value added is not available. As an alternative measure I then use the labor compensation share of the finance industry, i.e., the compensation of all employees of the finance industry divided by the compensation of all employees in the US economy.

Figure 4 displays various measures of the share of the finance and insurance industry in the GDP of the United States estimated from 1870 to 2012. For the period 1947–2012, I use value added and compensation measures from the Annual Industry Accounts of the United States, published by the Bureau of Economic Analysis (BEA). For the post-war period, the two measures display the same trends. This means that, in the long run, the labor share in the finance industry is roughly the same as the labor share in the rest of the economy (in the short run, of course, profit rates can vary). For 1929–1947, I use the share of employee compensation because value-added measures are either unavailable or unreliable. For 1870–1929, I use the Historical Statistics of the United States (Carter et al. 2006).\[9\]

There are three important points to take away from Figure 4. First, the finance income share varies a lot over time. Second, the measures are qualitatively and quantitatively consistent. It is thus possible to create one long series simply by appending the older data to the newer ones. Third, finance as a share of GDP was smaller in 1980 than in 1925. Given the outstanding real growth over this period, it means that finance size is not simply driven by income per capita.

B. Adjusted Measures

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

Wars.—During peace time and without structural change, it would make sense to simply use GDP as the relevant measure of total income. Two factors can complicate the analysis, however. First, WWI and WWII take resources away from the normal production of goods and services. Financial intermediation should then be compared to the non-war-related GDP. To do so, I construct a measure of GDP excluding

\[9\] Other measures based on Martin (1939) and Kuznets (1941) give similar values. More details regarding the various data sources can be found in Philippon and Reshef (2012) and in the online Appendix.
defense spending. The second issue is the decline in farming. Since modern finance is related to trade and industrial development, it is also useful to estimate the share of finance in non-farm GDP.

The left panel of Figure 5 presents the finance share of non-defense GDP, and of non-farm, non-defense GDP (or compensation, as explained above). Both adjustments make the series more stationary. In particular, using non-defense GDP removes the spurious temporary drop in the unadjusted series during WWII.

I use the defense-adjusted share as my main measure. The share of finance starts just below 2 percent in 1880. It reaches a first peak of almost 6 percent 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 percent of non-defense GDP.

**Other Services.**—Is finance different from other service industries? Yes. The right panel of Figure 5 also plots the share of finance in service GDP. It is mechanically higher than with total GDP, but the pattern is the same (the other fast growing service industry is health care, but it does not share the U-shaped evolution of finance from 1927 to 2009).

**Globalization.**—Figure 4 shows finance income divided by US GDP. This might not be appropriate if financial firms export some of their services abroad. It turns out, however, that globalization does not account for the evolution of the finance income share. There are two ways to show this point.

The right panel of Figure 5 displays the ratio of *domestic* finance income to (non-defense) GDP. Domestic income is defined as income minus net exports of financial services. The figure is almost identical to the previous one. The reason is that the United States, unlike the United Kingdom for instance, is not a large exporter of financial services. According to IMF statistics, in 2004, the UK financial services trade balance was +$37.4 billion while the US balance was −$2.3 billion: the United States was actually a net importer. In 2005, the UK balance was +$34.9 billion, and the US balance was +$1.1 billion. In all case, the adjustments are small.

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

### II. Quantity of Intermediated Assets

I measure the quantity of intermediated financial assets as follows:

\[
q_t \equiv b_{c,t} + m_t + k_t,
\]

where \(b_{c,t}\) is consumer credit outstanding, \(m_t\) are holdings of liquid assets, and \(k_t\) is the value of intermediated corporate assets (for the nonfinancial sector). The measurement principle is to take into account the instruments on the balance sheets of end users, households, and nonfinancial firms. This is the correct way to do the accounting, rather than looking at financial intermediaries’ balance sheets which reflects (in part) activities within finance itself.

Equation (2) is consistent with a model where it costs the same to extend $1 of consumer credit, $1 of business credit (or equity), or to create $1 of liquidity. That these costs are the same is far from obvious but is in fact consistent with
microeconomic evidence available for the more recent part of the sample. These assumptions as well as the underlying model are discussed in the online Appendix. I maintain for now the assumption that the relative costs of various types of intermediation remain constant over time and that the composition of borrowers remains constant. I relax this assumption in Section III. In the remainder of this section, I construct empirical proxies for $b_c$, $m$, and $k$.

A. Debt and Equity

Figure 7 presents credit liabilities of farms, households, and the business sector (corporate and non-corporate). These include all bank loans, consumer credit, mortgages, bonds, etc. The first point to takeaway 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 nonfinancial 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.10

10 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 nonfinancial issuers. What should count as output for the finance industry are only issuances by nonfinancial firms.
To extend the credit series before 1920, I use data on home mortgages provided by Schularick and Taylor (2012). I also use the balance sheets of financial firms. I measure assets on the balance sheets of commercial banks, mutual banks, savings and loans, federal reserve banks, brokers, and life insurance companies. I define total assets as the sum of assets of all these financial firms over GDP. I use this series to extend the total nonfinancial debt series (households and non-corporates, farms, corporates, government). I regress total credit on total assets and use the predicted value to extend the credit series.

The finance industry not only manages existing assets, but it also originates new assets and replaces old ones as they expire. It is therefore useful to consider stocks and flows separately. Figure 8 shows the issuances of corporate bonds by nonfinancial corporations as well as a measure of household credit flows. 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 8 also shows a measure of household debt issuance.

I use three measures of equity intermediation: total market value over GDP, initial public offering (IPO) proceeds over GDP, and gross (nonfinancial) equity offerings over GDP. Figure 9 shows that gross equity flows were high in the early part of

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

11 When I do not have a separate measure of flows, I assume a runoff rate consistent with the average ratio of flow to level, and I create the flow measure from the level series. Details are in the online Appendix.

12 Why use the market value of equity when thinking about intermediation? First, the rise in market value could be driven by improvements in financial intermediation. Two prime examples are risk management with financial derivatives (discussed in the online Appendix) and lower costs of participation in the equity market. Improvement in financial intermediation would lead to higher market value of equity. Clearly, in this case, the measure of unit cost
would be correct only if equity is measured at market value. Another reason for using market values is that book values miss a lot of intangible investments. On the other hand, changes in market values may reflect factors that are not directly related to financial intermediation, such as changes in household risk aversion or bubbles. This then begs the question of what is the “production function” of asset management services. The evidence in Greenwood and Scharfstein (2013) is consistent with a constant fee in the aggregate, even though individual fees might have decreased. Gennaioli, Shleifer, and Vishny (2014a) discuss this issue in details. Finally, notice that in the extended model of Section III, it is only the asset management fee that is proportional. The monitoring cost is not.
the sample. The market value of equity, on the other hand, is higher in the post-war period. The IPO series will allow me to implement quality-adjustments in the next section.

B. Money and Liquidity

In addition to credit (on the asset side of banks), households, firms, and local governments benefit from payment and liquidity services (on the liability side of banks and money market funds). For households, I use total currency and deposits, including money market fund shares, held by households and nonprofit organizations. The left panel of Figure 10 shows the evolution of this variable.

An important element to take into account in the measurement of liquidity provision is the rise of the shadow banking system. Gorton, Lewellen, and Metrick (2012) argue that a significant share recent activities in the financial sector was aimed at creating risk free assets with money-like features. For firms (incorporated or not), I follow Gorton, Lewellen, and Metrick (2012), and I treat repos as shadow deposits. The series is thus the sum of checkable deposits and currency, time and savings deposits, money markets mutual funds shares, and repos (by nonfinancial firms).13

C. Aggregation

If we could observe the income flows $y_{i,t}$, associated with the three fundamental sources of revenues $i = b, m, k$, we would simply compute the unit cost as, for instance: $\psi_{c,t} = \frac{y_{c,t}}{b_{c,t}}$, where $y_{c,t}$ would be the income generated by credit intermediation for consumers. Unfortunately, there is no satisfactory way to link a particular income to a particular activity, especially over long periods of time.14

---

13 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.

14 There is an empirical problem and a conceptual problem. Empirically, our data is organized by industry (e.g., securities, credit intermediation), not by function and even less by end-user. Even obtaining detailed measures of gross output is challenging. See Greenwood and Scharfstein (2013) for an enlightening discussion. But this is not only an issue of accounting. Even if we had all the data imaginable, we would still need to decide how to allocate...
This precludes a direct estimation of the $\psi_{i,t}s$. We only observe the total income of the finance industry, $y^f_t$, described in Section I. This is why I assume that the relative costs remain constant over time. The online Appendix shows how they can be estimated and that, in fact, they are close to one.

M&As.—An important activity of financial intermediaries is advising on mergers and acquisitions (M&As). 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 Securities Data Company (SDC) and Bloomberg to compute the value of merger deals. I then use historical data from Jovanovic and Rousseau (2005) to extend the series back to 1890. The next step is to apply the proper weight to the M&A series. M&A fees typically range from 1 percent for large deals to 4 percent for smaller ones. I assume that merger fees are 2 percent of the volume. This assumption is probably a bit higher than the weighted average fee, but there are also probably some ancillary activities associated with mergers and for which the finance industry is compensated.

**Flows and Stocks of Intermediated Assets.**—I construct two measures, one for the flow of new intermediation, one for the stock of outstanding intermediated assets.

Some activities are more naturally linked to flows (screening, IPO fees, etc.), and some are more naturally linked to stocks (debt restructuring, asset management, etc.). The stock measure is simply the sum of outstanding values

$$ q^\text{level}_t = b^\text{level}_{c,t} + b^\text{level}_{k,t} + e^\text{level}_{k,t} + m_t. $$

Note that $e^\text{level}_{k,t}$ is the market value of equity, as discussed earlier. For the flow measure, I also add up the values of new issuances, but I take into account the fact that underwriting fees are higher for equity than for debt (see Altinkiliç and Hansen 2000 and the online Appendix for details),

$$ q^\text{flow}_t = b^\text{flow}_{c,t} + b^\text{flow}_{k,t} + 3.5 e^\text{flow}_{k,t} + M^&A_t. $$

It corresponds to gross issuances of debt and equity, plus the value of mergers and acquisitions.\footnote{In theory, I would also need to take into account the debt of the government. The issue is which weight to apply. Government debt is risk-free and liquid, and it might actually help the functioning of financial markets and justify a negative weight (Krishnamurthy and Vissing-Jorgensen 2012; Greenwood, Hanson, and Stein 2011). But any long-term debt carries duration risk and positive intermediation costs. As a benchmark I set the weight to zero. The results are essentially unchanged if I set the weight to 1/10 instead.}

Note that the liquidity measure is only a level measure, and that the M&A measure is only a flow measure. Finally, the total measure of intermediated assets is

$$ q_t = q^\text{flow}_t + q^\text{level}_t. $$
The aggregate flow and stock measures \( (q_t^{\text{flow}}, q_t^{\text{level}}) \) are displayed in the left panel of Figure 11. The flow measure is an order of magnitude smaller than the stock measure. The flow measure collapses quickly during the Great Depression while the level measure peaks later and is exacerbated by deflation. A similar pattern emerges during the Great Recession. Overall, the stock measure increases more in recent years, driven by the market value of corporate equity and by the size of the household debt market.

The right panel of Figure 11 presents the total measures corresponding to four broad functions discussed earlier: credit and equity intermediation services to firms, credit intermediation services to households, liquidity services, and M&A activities. It is clear from Figure 11 that the intermediation series for firms and households are the most volatile ones. There is also a significant increase in liquidity services in the 2000s. M&As play some role mostly in the 1990s. By construction, the sum of the two series in the left panel of Figure 11 is the same as the sum of the four series in the right panel, and is equal to the measure of intermediated assets, \( q \), in Figure 2.

D. Evidence of Constant Returns to Scale

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

An important assumption of the model is that financial services are produced under constant returns to scale. Figure 12 presents evidence consistent with this assumption. It uses the period 1947–1973, for two reasons. First, the post-war data is the most reliable, and stopping in 1973 allows me to exclude major oil shocks, inflation, and other factors that might create short-term noise in my estimates. Second, as I will discuss shortly, quality adjustments are less important over this period than either before or after. Since these adjustments are difficult to implement, it is more convincing to first present the evidence without them.
From 1947 to 1973, real GDP per capita increases by 80 and real financial assets by 250 percent (measured in constant dollars), but my estimate of the unit cost of intermediation remains fairly constant (all series are presented as ratios to their values in 1950). By 1970 people are a lot richer, financial markets are a lot larger, but the unit cost is exactly the same as in 1950. This provides clear evidence that the production of financial services has constant returns to scale.

III. Quality Adjustments

The quantities of intermediation should be adjusted for the difficulty of monitoring/screening borrowers, otherwise the unit cost measure could register spurious changes in intermediation efficiency. These adjustments require a model. The model economy consists of households, a nonfinancial business sector, and a financial intermediation sector. In the model, the finance industry provides three types of services to households and firms: liquidity, monitoring, and asset management. Households hold the corporate capital stock via intermediaries. In addition, households borrow and lend from each other. The key point of the model is that households and firms are heterogeneous in their intermediation intensities. Some borrowers/issuers require more screening and monitoring that others.

A. Corporate Finance

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

Let us therefore consider a simple moral hazard model with heterogeneous firms. If a firm hires $n$ workers it produces $f(n)$ units of output, where $f$ is increasing and concave. Firms choose employment to maximize (detrended) net income $\pi(w) \equiv \max_n f(n) - wn$. There are two types of firms, $l$ and $h$, that differ in their cash on hand $x$ (equivalently in their retained earnings or their pledgeable collateral). I assume that $x^l < x^h$ and I refer to $l$-firms as low cash firms. There is an exogenous potential supply $k_h$ of $h$-firms and free entry of $l$-firms.\footnote{Let $k_t$ be the (endogenous) number of active firms, and let $n_t$ be employment per firm (so aggregate employment is $n_t = k_t n_t$). The number $k_t$ captures the extent to which investment opportunities occur in established companies. I assume that it is given by technology, and indeed, the data supports the view that large changes in $k_h$ are driven by large scale technological change (electricity, information technology). Note that the number of low cash firms $k_t - k_h$ is endogenous, and in particular, highly dependent on financial intermediation. So the way the model is going to interpret the 1990s is that established firms were not the ones able to promote the IT revolution. Instead it had to be younger firms, that are cash poor and therefore more dependent on financial intermediation. To the extent that we actually observe a large entry of young firms, the model will infer that financial intermediation must have been relatively efficient.}

To capture financial intermediation in a tractable way, I assume that capital can be diverted. The online Appendix describes the details of moral hazard and endogenous monitoring. The key point is that the model delivers the following monitoring demand function

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

where $\xi$ is the fraction of capital that can be diverted if there is no monitoring and $\varphi$ is a proportional intermediation cost, akin to asset management fees. The function $\mu(x)$ measures the quantity of intermediation services required for a firm with cash on hand $x$. Firms with high values of $x$ require less monitoring than firms with low values of $x$. The unit cost of monitoring is $\zeta_t$ and the income received by intermediaries for their monitoring activity is $\zeta_t \mu - t = \mu_h + (1 + r)(x_h - x_l) s_t$.

$$\bar{\mu}_t \equiv \mu_h + (1 + r)(x_h - x_l) s_t,$$

and $s_t \equiv \frac{k_l}{k_l + k_h}$ is the share of low cash firms in aggregate investment. The total income for corporate intermediation services is

$$y_{k,t}^f = \varphi k_t + \zeta_t \bar{\mu}_t.$$

Similarly, external finance (the quantity of monitored assets) is $\bar{b}_k = 1 - x^h + (x_h - x_l) s$. Note that the unit cost of external finance $\zeta_t \frac{\bar{\mu}_t}{\bar{b}_t}$ depends on the intensity of monitoring $\bar{\mu}/\bar{b}$, which changes with the share of low cash firms $s$. The parameter of interest is $\zeta_t$ which captures the true efficiency of financial intermediation. To recover $\zeta_t$, I need to estimate $\bar{\mu}/\bar{b}$.\footnote{Let $k_t$ be the (endogenous) number of active firms, and let $n_t$ be employment per firm (so aggregate employment is $n_t = k_t n_t$). The number $k_t$ captures the extent to which investment opportunities occur in established companies. I assume that it is given by technology, and indeed, the data supports the view that large changes in $k_h$ are driven by large scale technological change (electricity, information technology). Note that the number of low cash firms $k_t - k_h$ is endogenous, and in particular, highly dependent on financial intermediation. So the way the model is going to interpret the 1990s is that established firms were not the ones able to promote the IT revolution. Instead it had to be younger firms, that are cash poor and therefore more dependent on financial intermediation. To the extent that we actually observe a large entry of young firms, the model will infer that financial intermediation must have been relatively efficient.}
Philippon (2008) uses Compustat to construct an empirical proxy for $s_t$, namely the share of aggregate investment that is done by firms that must borrow more than three-quarters of their capital spending. The measure is displayed Figure 13. Following Eisfeldt and Rampini (2006), I have also computed measures of investment that include capital reallocation by adding acquisitions minus sales of used capital for each firm. All these measures are similar and suggest that the intensity of corporate finance was higher in the 1980s and 1990s than in the 1960s. Since these measures are based on Compustat data, they are available only from 1950 onward (at best). Figure 13 also shows IPO proceeds, based on the work of Jovanovic and Rousseau (2001) and Ritter (2011). The two series are highly correlated in the post-war period, and I use the IPO series to extrapolate the low cash share series before 1950, using a simple linear regression of one variable on the other. As argued by Jovanovic and Rousseau (2001), the IPO market of the 1920s was remarkably active, even compared to the one of the 1990s: IPO firms were of similar ages, and the proceeds (as share of GDP) were comparable.

### B. Household Finance

On a per-dollar basis, it is more expensive to lend to poor households than to wealthy ones, and relatively poor households have gained access to credit in recent years.\textsuperscript{17} To capture this idea, I assume that there is a continuum households and that there is a fixed cost to borrowing $\kappa$, in addition to the marginal cost $\varphi$. Income

\textsuperscript{17}Using the Survey of Consumer Finances, Moore and Palumbo (2010) document that between 1989 and 2007 the fraction of households with positive debt balances increases from 72 to 77 percent. This increase is concentrated at the bottom of the income distribution. For households in the 0–40 percentiles of income, the fraction with some debt outstanding goes from 53 to 61 percent between 1989 and 2007. In the mortgage market, Mayer and Pence
inequality among households is captured by the labor endowment $\eta$. The model is described in details in the online Appendix. The model features both an extensive margin (participation of households in the credit market) and an intensive margin (how much each household borrows). The extensive margin is characterized by the cutoff $\hat{\eta}$ such that only households with income above $\hat{\eta}$ use the credit market. The fraction of households (of a given generation) who have access to credit is therefore $1 - F(\hat{\eta})$ where $F$ is the c.d.f. of $\eta$.

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

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

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

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

### C. Calibrated Model

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

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

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

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

Figure 14 studies the impact of changes in the unit cost of asset management $\Phi$, which is calibrated to 1 percent in 1989. An increase in the cost of asset management increases the finance share of GDP (panel A), and decreases the size of the credit market (panel B), as expected. Note that the model, unlike the data, allows me to separate the income received from corporate finance and from household finance. The question is whether the unit costs $\Psi_k = y^f_k/k$ and $\Psi_c = y^f_c/b_c$ correctly capture the changes in $\Phi$. Panel C shows that the answer is “almost.”

Panel A presents separately the income from corporate finance services and the income from household finance services. Panel B shows firm’s external finance and household debt. Panel C shows the adjustment to the particular class of intermediated assets needed to remove the bias in the measurement of intermediation costs. When the adjustment is above 1, the unit cost is overestimated and the quantity of assets must be scaled up to obtain the correct estimate.

Notes: Comparative statics using the calibrated model (Table 4 of online Appendix). The horizontal axis is $\Phi/\Phi_0$ where $\Phi_0$ is calibrated to 1 percent in 1989. Panel A presents separately the income from corporate finance services and from household finance services. Panel B shows firm’s external finance and household debt. Panel C shows the adjustment to the particular class of intermediated assets needed to remove the bias in the measurement of intermediation costs. When the adjustment is above 1, the unit cost is overestimated and the quantity of assets must be scaled up to obtain the correct estimate.

Figure 14 studies the impact of changes in the unit cost of asset management $\Phi$, which is calibrated to 1 percent in 1989. An increase in the cost of asset management increases the finance share of GDP (panel A), and decreases the size of the credit market (panel B), as expected. Note that the model, unlike the data, allows me to separate the income received from corporate finance and from household finance. The question is whether the unit costs $\Psi_k = y^f_k/k$ and $\Psi_c = y^f_c/b_c$ correctly capture the changes in $\Phi$. Panel C shows that the answer is “almost.” Even for very large...
changes in $\varphi$ (from 0 to twice the benchmark value), the bias barely exceeds 5 percent. In a sense, this is not surprising because we are only changing the linear part of the model in this experiment. When $\varphi$ goes down, more potential borrowers actually borrow, and each borrower borrows more. This does not create a significant bias.

Figure 15, on the other hand, studies the highly nonlinear part of the model, by changing the share of low cash firms and the fixed cost of participation for households. Panels A, B, and C focus on changes in the composition of firms. The exogenous forcing variable is the number of cash-rich firms $k_h$. Note that the figure shows the response of the various variables as a function of the observed share of low cash firms, which is itself an endogenous variable. The reason is that $s$ is observable in the data and will be used to make the quality adjustment. The benchmark model is calibrated using a share of low cash firms of 20 percent (in 1989). When this

the unit cost. To adjust this measure, I define $\tilde{\gamma}^I$ as counterfactual income if we wanted to obtain $Q$ with the initial technology $\zeta_0$, in other words $\tilde{\gamma}^I = Y(\chi_0, Q)$. The conceptually correct change in the unit cost is $\gamma^I/\tilde{\gamma}^I_0$ since by construction $Q$ is unchanged. If $\psi/\psi_0 > \gamma^I/\tilde{\gamma}^I_0$, then the empirical measure overestimates the change in unit cost. Then the adjustment is defined as $\psi/\psi_0 \gamma^I$. This adjustment has the property that if I use it to artificially scale up $q$, I recover the correct value for $\psi/\psi_0 = \gamma^I/\tilde{\gamma}^I_0$. This adjustment can be applied to the entire amount of intermediated assets, or to particular classes, such as $b_h, b_k$, etc.
share increases, monitoring costs and external finance both increase (panels A, B). Monitoring intensity increases, and this creates a measurement bias in the sense that the perceived unit cost increases. If the share is 40 percent, the model says that external corporate finance should be scaled up by roughly 25 percent in order to remove the induced bias in the measurement of the unit cost.

Panels D, E, and F in Figure 15 focus on changes in the availability of household credit induced by exogenous changes in the fixed cost $\kappa$. When $\kappa$ increases, some relatively poor households are priced out and participation in the credit market falls (D). The model is calibrated to a participation of 84 percent and a household debt to GDP ratio of 73 percent in 1989. If $\kappa$ doubles, the participation rate and the debt/GDP ratio drop to approximately 60 percent. The participation rate drops more than the debt/GDP ratio because rich households still borrow, and they typically borrow more (when young) that poor households. These nonlinear composition effects create again a significant bias.

D. Adjusted Unit Cost

The goal of this section is to use the calibrated model presented above to adjust the asset series of Figures 11. The first step is to choose which adjustments to make. I take away from Figure 14 that changes in the proportional cost $\varphi$ are unlikely to create significant biases. I therefore focus on the other parameters. At the firm level, the choice is fairly obvious: Figure 13 shows that the share of low cash firms changes a lot over time, and panel C of Figure 15 shows that this can create large biases. The implementation is straightforward: plug in the observed value of $s$, read the adjustment factor in panel C of Figure 15, and multiply the empirical series for $b_k$ by this factor. The implied series is “Firm adj.” in Figure 16. As expected, the adjustment is quantitatively important in the 1920s and in the 1980s and 1990s, which correspond to waves of innovation driven by new technologies.

Biases in the household debt market are likely to come from changes in household participation. The adjustment is difficult because I do not have a long time series for the participation rate of households. Since the goal of this section is to assess measurement biases, I will look for the maximal adjustment by assuming that changes in the household debt to GDP ratios are driven by changes in the fixed cost $\kappa$.

The series “Firm and HH adj.” in Figure 16 shows the output measures with quality adjustments for both corporate and household finance. Adjustments to consumer credit matters mostly after 1970. As argued earlier the quality adjustments are small between 1947 and 1973, which makes it an ideal period to test the constant returns to scale assumption. The adjustment is large in the recent part of the sample. After

There is prima facie evidence of technological change in the intermediation technology (e.g., credit scoring) that has made it easier for poor households to obtain credit. So we know that this account for some of the evolution of the household debt market, but we do not know precisely how much. I am going to interpret the historical time series as if the growth in consumer credit mostly reflects improvements in intermediation. I only impose the constraint that the predicted participation rate cannot exceed 100 percent. This constraint binds in the model in the years 2000s, which is consistent with the view that household debt growth was linked to house prices for households who already had access to credit. I have also considered the implications of changes in inequality, but I have found that these are unlikely to create significant biases. Changes in inequality typically change the debt/GDP ratio and the finance income share, but the quantitative experiments suggest that the unit cost is not severely biased.
1990 the unadjusted measure of business intermediation underestimates intermediation by about 25 percent.

Table 1 and Figures 16 and 17 are the main contributions of the paper. They bring together the historical/empirical work of Sections I and II, and the theoretical/quantitative work of Section III (an adjustment for non-life insurance services is discussed below). Figure 17 shows the unit cost of financial intermediation, defined as income divided by adjusted intermediated assets. There are two main points to take away. The first is that the unit cost ratio is remarkably stable. Recall that we start from series—for income, debt, equity, etc.—that fluctuate a lot over time. But their ratio is stable. The simple unit cost series has a mean of 1.87 percent and a volatility of 23 basis points. Quality adjustments increase the volatility of the assets series but reduce the volatility of the unit cost measure, by about 25 percent. The adjusted series has a standard deviation of only 16 basis points. The second main point is that the unit cost of financial intermediation is about the same today as it was around 1960 and 1900.

E. Insurance Services

The model is designed to account for consumption smoothing that takes place via credit markets. In the model, an improvement in household finance leads to more borrowing and better consumption smoothing.20 Insurance companies, however,

\footnote{For instance, Gerardi, Rosen, and Willen (2010) find that the purchase price of a households home predicts its future income. The link is stronger after 1985, which coincides with important innovations in the mortgage market. The increase in the relationship is more pronounced for households more likely to be credit constrained. The model}
provide services that are not directly related to intermediation. This is potentially an issue since the income of insurance companies is counted as a cost of intermediation, while the services provided might not be well captured by standard measures of intermediated assets. I therefore attempt a (rough) adjustment by subtracting consumption expenditures on non-life insurance services (health insurance, household insurance, motor vehicle, and other transportation insurance) from the total income of intermediaries. The quantitative significance of this adjustment comes from motor vehicle insurance which grows rapidly in the 1950s and is around one-half captures correctly measures these effects, and consumption smoothing that entails the creation of credit flows does not create a bias in my estimation. Informal risk sharing, for instance within families, would be enter neither the income side, not the asset side of my calculations, so it should not create a bias either. The overall evidence on risk sharing is mixed. Income inequality has increased dramatically in the United States over the past 30 years. If financial markets improve risk sharing, however, one would expect consumption inequality to increase less than income inequality. This is a controversial issue, but Aguiar and Bils (2011) find that consumption inequality closely tracks income inequality over the period 1980–2007. Therefore it seems difficult to argue that risk sharing among households has improved significantly over time.

Table 1—Estimated Financial Intermediation

<table>
<thead>
<tr>
<th>Variable</th>
<th>Observations</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finance income/GDP</td>
<td>127</td>
<td>0.0422</td>
<td>0.0169</td>
<td>0.0194</td>
<td>0.077</td>
</tr>
<tr>
<td>Finance income/GDP w/o ins.</td>
<td>127</td>
<td>0.0359</td>
<td>0.0132</td>
<td>0.0163</td>
<td>0.063</td>
</tr>
<tr>
<td>Intermediated assets/GDP</td>
<td>127</td>
<td>2.241</td>
<td>0.726</td>
<td>1.046</td>
<td>3.90</td>
</tr>
<tr>
<td>Inter. assets, firm adj.</td>
<td>127</td>
<td>2.369</td>
<td>0.816</td>
<td>1.062</td>
<td>4.17</td>
</tr>
<tr>
<td>Inter. assets, firm and HH adj.</td>
<td>127</td>
<td>2.511</td>
<td>0.943</td>
<td>1.091</td>
<td>4.70</td>
</tr>
<tr>
<td>Unit cost</td>
<td>127</td>
<td>0.0187</td>
<td>0.0022</td>
<td>0.0145</td>
<td>0.0235</td>
</tr>
<tr>
<td>Unit cost, firm adj.</td>
<td>127</td>
<td>0.0178</td>
<td>0.0017</td>
<td>0.0140</td>
<td>0.0222</td>
</tr>
<tr>
<td>Unit cost, firm and HH adj.</td>
<td>127</td>
<td>0.0169</td>
<td>0.0016</td>
<td>0.0132</td>
<td>0.0217</td>
</tr>
<tr>
<td>Unit cost, adj. w/o ins.</td>
<td>127</td>
<td>0.0146</td>
<td>0.0020</td>
<td>0.0104</td>
<td>0.0204</td>
</tr>
</tbody>
</table>

Note: Data range 1886–2012.

Figure 17. Quality Adjusted Unit Cost of Intermediation

Note: Total intermediation costs divided by quality-adjusted intermediated assets.
of 1 percent of GDP today, and health insurance which grows linearly to reach about 1 percent of GDP. Whether or not these services ought to be included in financial intermediation is debatable. On the one hand, these services differ significantly from banking and traditional intermediation services. On the other hand, they are financial services linked to the consumption of particular goods, and they certainly affect precautionary savings decisions and therefore the size of the credit market. Removing all of these expenditures is probably an over-adjustment, so the unit cost without (non-life) insurance in Figure 17 should be seen as a lower bound on the true unit cost. The new series suggests a slight downward trend in unit cost until 1970. It does not change the main point regarding the post-war sample: the unit cost is still low in the 1960s, and the discrepancy with the 2000s is at least as large as before.

F. Discussion of the Results

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

A potential explanation is oligopolistic competition but the link between market power and the unit cost of intermediation is not easy to establish. Adding a constant markup of price over marginal cost would not change anything to the trends presented above. The issue is whether market power changes over time. The historical evidence does not seem to support the naïve market power explanation, however. Regulatory barriers to entry have been reduced in banking since the 1970s and yet this is when the unit cost goes up. Conversely, if there is one period where we have a strong presumption that banks had significant market power, this must be the turn of the nineteenth century. But the late Gilded Age is not a period where the unit cost of intermediation is high. The link between market power and the unit cost is therefore rather tenuous and more research is needed on this important topic.

---

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

22 Why do people trade so much? Financial economics does not appear to have a good explanation yet. One explanation is overconfidence, as in Odean (1998). Recent work by Glode, Green, and Lowery (2012) and Bolton, Santos, and Scheinkman (2011) explains why some type of informed trading might be excessive. Pagnotta and Philippon (2011) present a model where trading speed can be excessive. In these models, advances in IT do not necessarily improve the efficiency of financial markets.
Another plausible explanation is that my measures might fail to capture the social value of information production in financial markets. This effect is elusive because it can show up as an improvement in TFP with little impact on the aggregate quantity of assets. The only way to test the information production hypothesis is then to estimate directly the informational content of asset prices, as Bai, Philippon, and Savov (2011) attempt to do. This is another area where more research is needed.

IV. Conclusion

I have provided benchmark measures for the aggregate income of the US finance industry, the quantity of intermediated assets, and the unit cost of financial intermediation. The income of the finance industry as a share of GDP fluctuates a lot over time. These fluctuations are mostly driven by equally large fluctuations in the quantity and quality of intermediated assets. The unit cost of financial intermediation represents an annual spread of 1.87 percent on average. The unit cost of intermediation does not seem to have decreased significantly in recent years, despite advances in information technology and despite changes in the organization of the finance industry. The methodology developed in this paper can be used to quantify these evolutions, as well as to compare the cost of financial intermediation across different countries.

REFERENCES


