# Finance vs. Wal-Mart: Why are Financial Services so Expensive?

Thomas Philippon, New York University

Abstract.
Despite its fast computers and credit derivatives, the current financial system does not seem better at transferring funds from savers to borrowers than the financial system of 1910.

*"I would rather see Finance less proud and Industry more content."*
Winston Churchill, 1925

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. Total compensation of financial intermediaries (profits, wages, salary and bonuses) as a fraction of GDP is at an all-time high, around 9% of GDP.

What does society get in return? Or, in other words, what does the finance industry produce? I measure the output of the finance industry by looking at all issuances of bonds, loans, stocks (IPOs, SEOs), as well as liquidity services to firms and households. Measured output of the financial sector is indeed higher than it has been in much of the past. But, unlike the income earned by the sector, it is not unprecedentedly high.

Historically, the unit cost of intermediation has been somewhere between 1.3% and 2.3% of assets. However, this unit cost has been trending upward since 1970 and is now significantly higher than in the past. In other words, the finance industry of 1900 was just as able as the finance industry of 2010 to produce loans, bonds and stocks, and it was certainly doing it more cheaply. This is counter-intuitive, to say the least. How is it possible for today's finance industry not to be significantly more efficient than the finance industry of John Pierpont Morgan?

What happened? Why did we get the bloated finance industry of today instead of the lean and efficient Wal-Mart? 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 US, the cost of intermediation appears to have

increased. So why is the non-financial sector transferring so much income to the financial sector?

One simple answer is that technological improvements in finance have mostly been used to increase secondary market activities, i.e., trading. Trading activities are many times larger than at any time in previous history. Trading costs have decreased, but I find no evidence that increased liquidity has led to better (i.e., more informative) prices or to more insurance.

## Measuring the Cost of Financial Intermediation

The sum of all profits and wages paid to financial intermediaries represents the cost of financial intermediation.

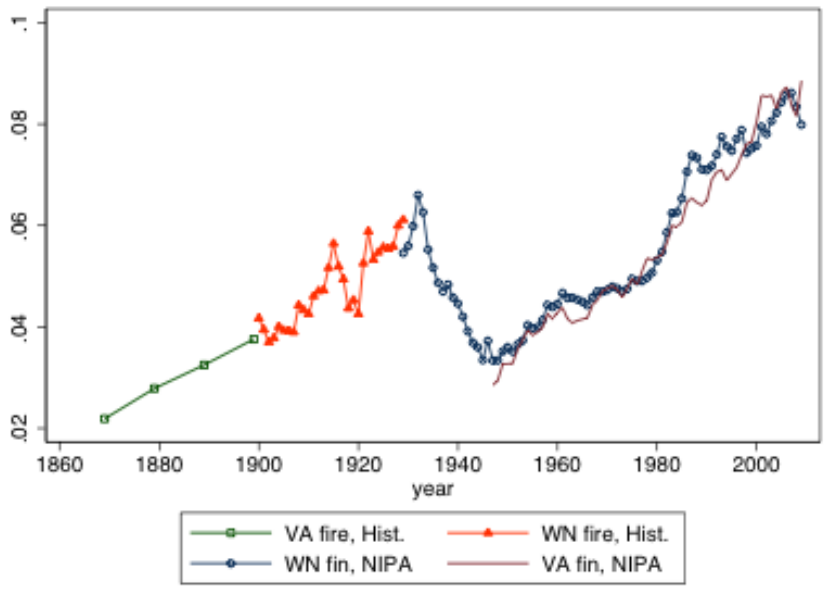
There are various ways to define the size of the financial sector. Conceptually, the measure is:

$$\text{Cost} = \text{Income of Finance Industry} / \text{Total Income}$$

The three most important issues are:

1. 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 homeowners. Whenever possible, I exclude real estate. In my notations, all variables indexed with "fin" include finance and insurance and exclude real estate.
2. Definition of "Income." The best conceptual measure is Value Added. In this case, "Cost" is GDP of the finance industry over the GDP of the US economy. However, this is only acceptable if we can exclude real estate, or at least imputed rents. When this is not possible, a good alternative is to use the compensation of employees. In this case, "Cost" 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 of the finance industry is the same as the labor share of the rest of the economy. In the short run, of course, profit rates can vary.
3. Definition of "Total Income." During peacetime and without structural change, it would make sense to simply use GDP. 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. This adjustment makes the series more stationary.

Figure 1: GDP Share of Finance Industry




I measure this cost from 1870 to 2010, as a share of GDP, and find large historical variations, shown in Figure 1 (with the various data sources, see Philippon 2011 for details).
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.

The cost of intermediation 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 almost 9% in 2010.

This pattern is not driven by globalization or by structural changes in the economy. The pattern remains the same if finance is measured as a share of services, and if net financial exports are excluded (see Philippon, 2011).

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 economic development.

## Measuring the Output of Financial Intermediation.

Next comes the issue of measuring the output of the financial sector. Following Merton (1995) and Levine (2005), one can propose the following four categories of financial services or functions:

- Provide means of payment (ease the exchange of goods and services)
- Produce information about investment opportunities

- Monitor investments and exert corporate governance
- Provide markets for insurance (diversification, risk management, liquidity)

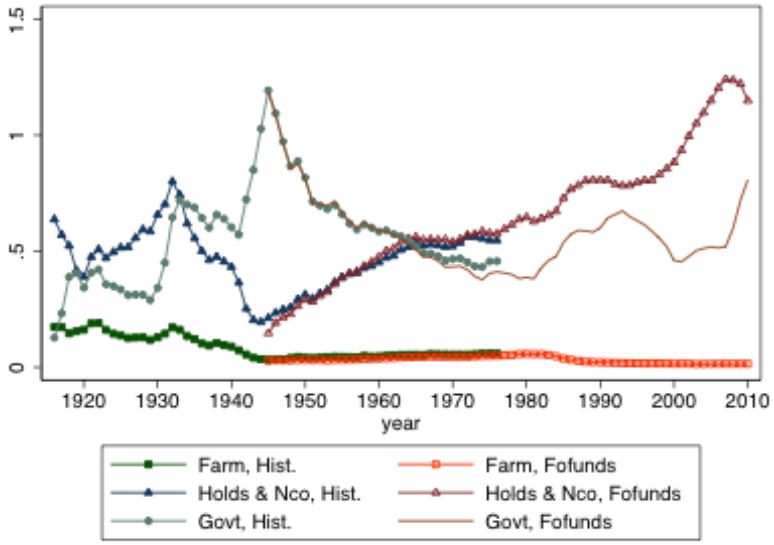
These services are the output of the finance industry and its source of economic value. To the extent that this higher total cost is met with proportionally more output, the greater compensation of the sector should not be surprising.

These services are provided to both households and firms, and facilitate the creation of financial assets. The most important contracts involve the credit markets. I measure the production of credit separately for households, farms, non-financial corporate firms, financial firms, and the government.

I show in Philippon (2011) that a simple benchmark can be constructed using the workhorse of modern macroeconomics, i.e., neo-classical growth model. This benchmark is a weighted average of the financial assets created by the financial sector for the real economy.

The most important trends in recent years are the increase in household debt, and in financial firms' debt. Figure 2 shows the outstanding bonds issued by Farms, Households and Non Profit Organizations, and the Government. Household debt exceeds 100% of GDP for the first time in history (see Figure 2), while financial debt exceeds non-financial corporate debt for the first time. Surprisingly, the non-financial corporate credit market is smaller today than it was at its peak of the late 1920s.

**Figure 2: Debt over GDP (selected sectors)**

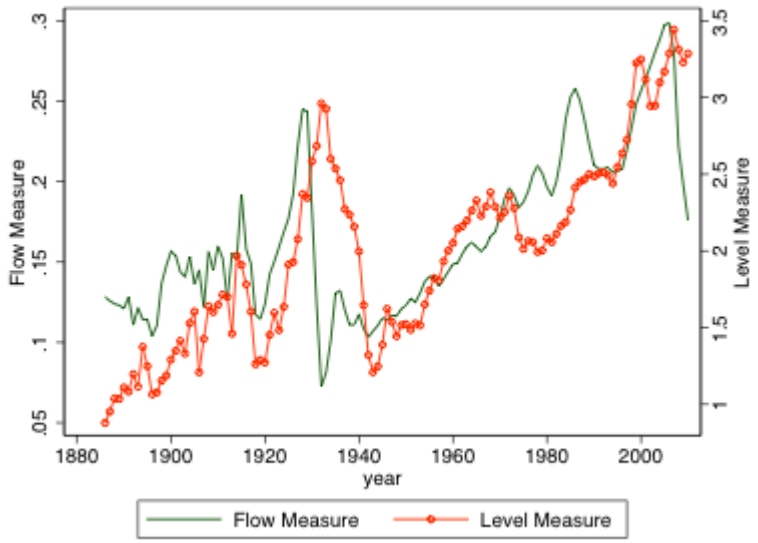



For the corporate sector, we need to look at bonds and stocks, and for stocks, we want to distinguish seasoned offerings and IPOs. We also need to look at the

liquidity benefits of deposits and money market funds. When we put all the pieces together, we obtain a series for output for the finance industry.

I then aggregate all types of non-financial credit, stock issuance, and liquidity services from deposits and money market funds.

**Figure 3: Financial Intermediation Output**




I construct two series of output in Figure 3: One using the flows (gross issuances over GDP) and one using the levels (debt 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.

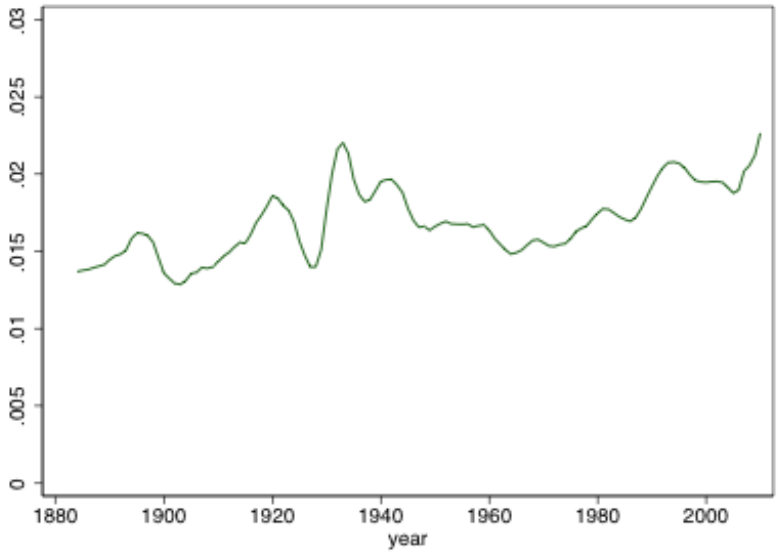
The two series are displayed in Figure 3. The production of financial services increases steadily until WWI, and rapidly after 1919 until 1929. It collapses during the great depression and WWII. It increases steadily until 1975 and more randomly afterwards. The flow and level measures share the same long term trends, but there are clear differences at medium frequencies. The flow variable is more stationary before WWI, suggesting a steady buildup of financial assets. The flow variable collapses much faster during the great depression and the great recession. The level variable peaks in 1933 because of deflation and the need to deal with rising default rates.

## The Decreasing Efficiency of Intermediation in the U.S.

I can then estimate the cost of financial intermediation, defined as the value added share divided by output series. The cost of intermediation in the US (expressed as a share of outstanding assets) is between 1.3% and 2.3%.

However, the cost of intermediation per dollar of assets created has increased over the past 130 years, and especially since the 1970s. In other words, according to this measure, the finance industry that sustained the expansion of railroads, steel and chemical industries, and the electricity and automobile revolutions was more efficient than the current finance industry.

**Figure 4: Financial Intermediation Unit Cost**




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?

It is important to understand that using the GDP share of finance to measure the costs of financial intermediation captures all fees and spreads but it ignores hidden costs of systemic risk. The neo-classical benchmark also assumes that all agents borrow rationally ex-ante (of course, that does not rule out the fact that they might end up with too much debt ex-post, but that means that they understand the risks involved and choose to borrow). We can debate this assumption, but the point I want to emphasize is that this provides an upper bound on financial efficiency. If anything, adding excessive risk taking and over-borrowing would decrease the risk-adjusted efficiency.[1]

[1] For an insightful discussion, see Haldane and Madouros (2011), Popov and Smets (2011), and Arcand, Berkes, and Panizza (2011).

## Information Technologies: Where is Wal-Mart When We Need It?

An obvious driving force in financial intermediation is information technology. I often hear the argument that improvement in IT explains the increase in the share of finance. This argument, however, is either incomplete or misleading. One reason it is incomplete is simply that IT cannot explain the evolution of the GDP share of finance before 1970.

What makes the IT argument misleading is that it is far from clear why IT should increase the share of finance. The neo-classical growth model predicts that, in most cases, technological improvement should lower the share of GDP spent on financial intermediation. In particular, this prediction is unambiguous for most retail finance. Essentially, the physical transaction costs of buying and holding financial assets must have decreased because of IT. This effect must have lowered the amount spent on intermediation.

An apt analogy is with retail and wholesale trade, since these are also intermediation services.[2] As Blanchard (2003) explains in his discussion of Basu et al. (2003), "fully one-third of the increase in TFP growth from the first to the second half of the 1990s in the United States came from the retail trade sector. For this reason, the general merchandising segment, which represents 20% of sales in the sector, was one of the sectors examined in a McKinsey study (McKinsey Global Institute, 2001) aimed at understanding the factors behind U.S. TFP growth in the 1990s."

Figure 5(a,b) shows the evolution of GDP shares and IT investment in wholesale trade, retail trade.

[2] For instance, one can compare retail finance and retail trade. Households go to grocery stores not because they derive utility from doing so, but rather to have access to groceries. Similarly, households use financial intermediaries to gain access to the financial products that they need.

**Figure 5 IT Investment and GDP Shares of Retail and Wholesale Trade**

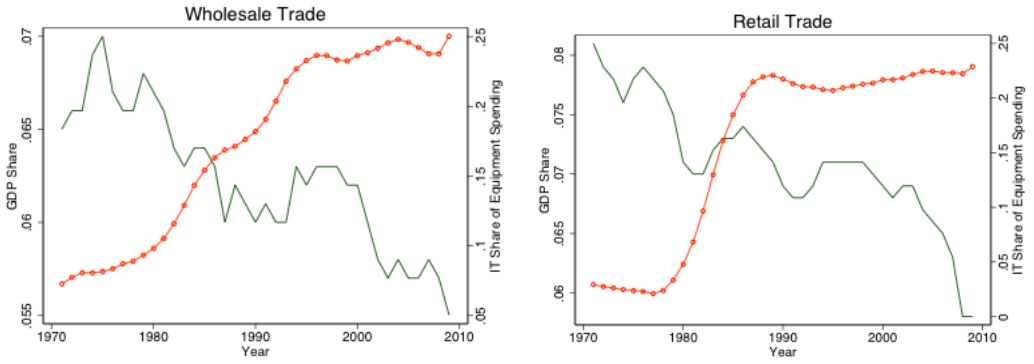
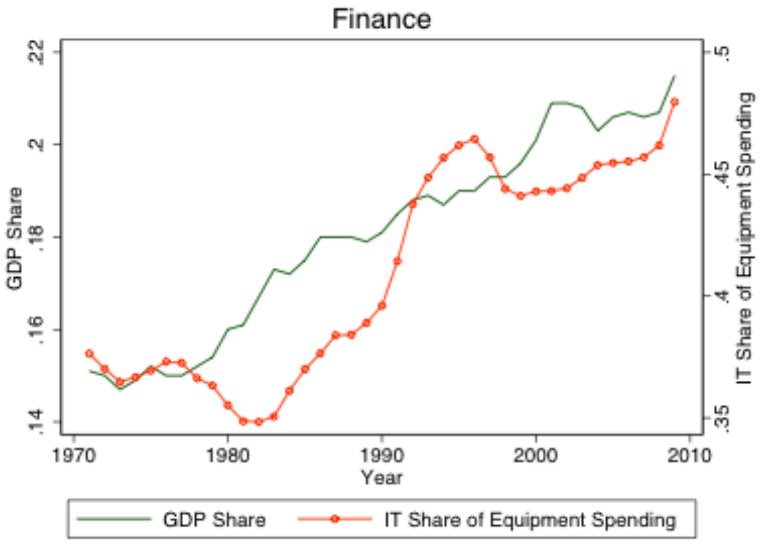



Figure 6 shows this evolution for finance. The contrast is striking. Based on what we see in wholesale and retail trade, IT should have made finance smaller, not larger.

**Figure 6: IT and GDP Share in Finance**



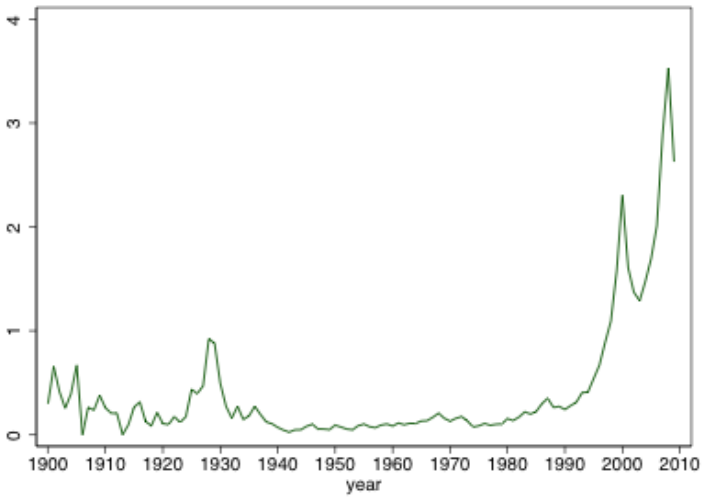

## Trading

What happened? Why did we get the bloated finance industry of today instead of the lean and efficient Wal-Mart? 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 US, the cost of intermediation appears to have increased. So why is the non-financial sector transferring so much income to the financial sector?

One simple answer is that technological improvements in finance have mostly been used to increase secondary market activities, i.e., trading. Figure 7 shows trading in the stock market. Foreign exchange trading volumes are more than 200 times larger today than in 1977. Finally, trading accounts for a large fraction of revenues for the largest banks.

Trading, of course, is neither a good nor a bad thing. It all depends on its impact on the real economy. The output measures developed above, however, only capture the production of financial assets (equity, bonds, money, etc.).[3] Two important functions of financial markets are not captured: the production of price information, and the provision of insurance.

**Figure 7: Equity Trading Volume over GDP**




It is then 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

In a model where managers learn from prices, better prices should lead to better capital allocation and higher productivity.

Are prices more informative about future income streams? Preliminary evidence in Bai, Philippon and Savov (2012) suggests no. They use a large panel of stock price data and ask whether a stock's price (relative to it's assets) contains more

[3] Note that the impact on average user costs is already taken into account. If trading lowers borrowing costs, firms can borrow more and invest more. This would be captured by the previous measures.

information about future earnings growth today than it has in the past. The answer from this statistical analysis is no.

## Risk Sharing

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. I am not aware, however, of any evidence suggesting improvement in risk management. The most obvious index, that of precautionary savings by businesses, suggests 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 a small fraction of all trades in the CDS market.

At the household level, better risk sharing should lead to less consumption risk. Income inequality has increased dramatically in the US 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 the vast sums of money spent on intermediation are justified by better risk sharing among households. It is also unclear that any of the main financial innovations of the past 20 years have improved risk sharing opportunities.

One area where there is evidence of improved consumption smoothing is in the housing market. Gerardi et al. (2010) find that the purchase price of a household's home predicts its future income. The link is stronger after 1985, which coincides with important innovations in the mortgage market. The increase in the relationship is more pronounced for households more likely to be credit constrained. My model, however, captures this type of smoothing and is reflected as increased mortgage borrowing. Therefore, it does not bias my estimates.

## Financial Derivatives

Derivatives markets have grown enormously. As of June 2011, the notional amount of outstanding OTC contracts is $700 Trillions (BIS, December 2011), with interest rates contracts (mostly swaps) accounting for $550T, and CDS for 32T. Measured at gross market values the numbers are 20T, 13T, and 1.3T, respectively.

These numbers are certainly impressive, but the relevant question is: do they bear any connection to measures of market and economic efficiency? The short answer is: no.

Most people are struck when they hear the market for financial derivatives is 700 Trillion. These numbers are sometimes used to justify the costs of financial services. This is misleading. Derivatives are just the plumbing of the financial markets. End-users do not care about the plumbing, only about the quality of the service.

Another analogy comes to mind. Most people would probably be struck if they heard that each Airbus A380 contains 40,300 connectors and 100,000 wires with a total length of 330 miles (530km). The wires in a single airplane would be enough to go from Philadelphia to Boston, or from Paris to London or Frankfurt. Should we congratulate Airbus when it manages to increase the length of its wires? Or should we only care about the safety, comfort, speed and fuel economy of the plane?

Similarly, suppose you were told that the "complexity and interconnectedness" inside your computer had increased. Would you pay for the pleasure to have a complex and interconnected computer, or would you only care about its speed, design and battery life?

To understand better the relevance (or lack thereof) of financial derivatives, consider the following example. Corporation A needs a long-term fixed-interest loan. Making the loan would expose the lender(s) to duration risk and to credit risk. How these risks are allocated, however, depends on the internal organization of the finance industry. Consider two polar cases. Suppose first that the loan is made and retained by bank B. Bank B must be compensated for bearing duration and credit risks. For instance, B must monitor its credit exposure and maintain a buffer of equity against credit risk. B must also monitor and hedge its interest rate risk. These activities are costly and the costs are passed through to the borrowers through spreads and fees.

Assume now that B can transfer credit risk to fund C using a CDS. Fund C now bears the credit risk, while B retains the duration risk. B and C must be compensated accordingly. The key point is that B and C together hold exactly the same risk as B in the earlier example. Absent other frictions, the two examples are exactly equivalent in terms of economic efficiency. Comparing the two polar cases, one can see that the size of the CDS market bears no connection to any measure of efficiency.

Let us now extend the example. 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. 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. With free entry, the total income going to intermediaries {B+C} would decrease. The unit cost measure developed earlier would correctly capture these effects: 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.

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 directly measure the informativeness of prices, or the welfare gains from risk sharing among non-financial firms and households. As explained earlier, however, I am not aware of any evidence suggesting better risk sharing or better prices.

## Conclusion

The finance industry of 1900 was just as able as the finance industry of 2000 to produce bonds and stocks, and it was certainly doing it more cheaply. But the recent levels of trading activities are at least three times larger than at any time in previous history. Trading costs have decreased (Hasbrouck (2009)), but 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.

In the absence of evidence that increased trading led to either better prices or better risk sharing, we would have to conclude that the finance industry's share of GDP is about 2 percentage points higher than it needs to be and this would represent an annual misallocation of resources of about $280 billions for the U.S. alone.

## References


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

Arcand, J.-L., E. Berkes, U. Panizza (2011), "Too much Finance?", VoxEU.org, 7 April.

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

Basu, S., J. G. Fernald, N. Oulton, and S. Srinivasan (2003): “The Case of the Missing Productivity Growth,” in NBER Macroeconomics Annual, ed. by M. Gertler, and K. Rogoff, vol. 18, pp. 9–63.

Blanchard, O. (2003): “Comment on Basu et al.,” in NBER Macroeconomics Annual

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

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

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.

Haldane, A.G. and V. Madouros (2011), "What is the Contribution of the Financial Sector, VoxEU.org, 22 November.

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

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

McKinsey Global Institute (2001), "US *productivity growth 1995-2000.* Understanding the Contribution of Information Technology Relative to other Factors. Washington, D.C.

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

Philippon, T. (2011) “Has the U.S. Finance Industry Become Less Efficient?” mimeo NYU.

Popov, A. and F. Smets (2011), “On the tradeoff between growth and stability: The role of financial markets”, VoxEU.org, 3 November.