FINANCE AND THE PRESERVATION OF WEALTH*

Nicola Gennaioli
Andrei Shleifer
Robert Vishny

We introduce the model of asset management developed in Gennaioli, Shleifer, and Vishny ("Money Doctors," *Journal of Finance*, forthcoming 2015) into a Solow-style neoclassical growth model with diminishing returns to capital. Savers rely on trusted intermediaries to manage their wealth (claims on capital stock), who can charge fees above costs to trusting investors. In this model, the ratio of financial income to GDP increases with the ratio of aggregate wealth to GDP. Both rise along the convergence path to steady state growth. We examine several further implications of the model for management fees, unit costs of finance, and the consequences of shocks to trust and to the capital stock. *JEL Codes*: D91, E21, E44, G11.

I. Introduction

Philippon (2013) documents the astonishing rise of the share of GDP coming from the financial sector since World War II (Figure I). Financial income rose from about 2% of the total in the 1940s to close to 8% at the time of the financial crisis. Philippon and Reshef (2013) document similar trends in many other developed countries. Greenwood and Scharfstein (2013) show that at least in the past 30 years, much of this rise of finance in the United States comes from financial services to consumers, especially asset management and credit intermediation of mortgages and consumer loans.

The growth of the financial sector has proved difficult to explain. Perhaps productivity in finance, as in other services, does not grow as fast as that in other sectors, so we see a manifestation of the Baumol (1967) disease. However, finance has grown relative to other services (Philippon and Reshef 2013), and wages in

*We are grateful to Robert Barro, Lawrence Katz, Yueran Ma, George Pennacchi, Thomas Philippon, David Scharfstein, Neil Thakral, and especially Charles-Henri Weymuller for extremely helpful comments, and to the Smith Richardson Foundation for financial support. Gennaioli thanks the European Research Council for financial support through the grant ERC- GA n. 241114. Shleifer was a co-founder of LSV Asset Management, a money management firm, but is no longer a shareholder in the firm. Shleifer's wife is a partner in a hedge fund, Bracebridge Capital. Vishny was a co-founder of LSV Asset Management. He retains an ownership interest.

The Quarterly Journal of Economics (2014), 1221–1254. doi:10.1093/qje/qju012.
Advance Access publication on May 7, 2014.

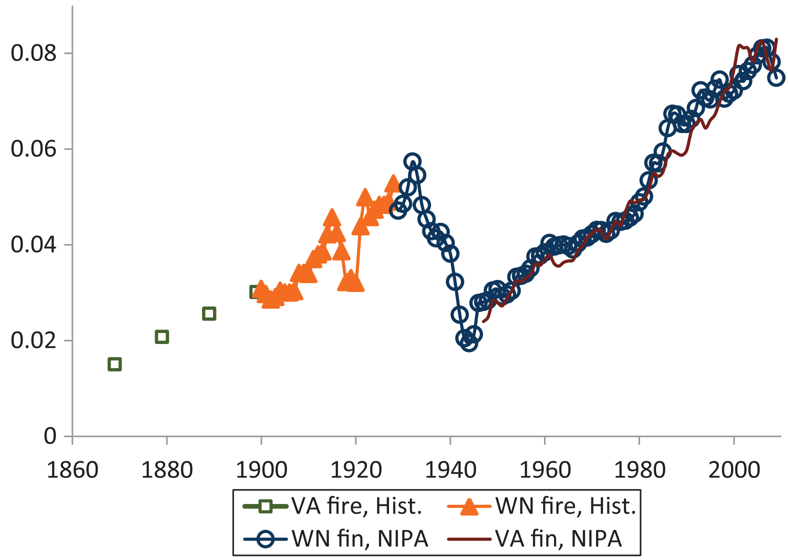

FIGURE I

Financial Sector Income/GDP

Source: Philippon (2013).

finance have grown faster than those in other service sectors (Philippon and Reshef 2012), inconsistent with this view. Philippon (2013) treats the cost of finance as a share of intermediated wealth due to screening and monitoring, but does not explain what determines this share or why financial income rises with market wealth. We present a new model of how financial income is endogenously determined as a function of intermediated wealth, describe what shapes this function, and explain how wealth and financial income move together.

Ours is a Solow-style growth model with a financial sector delivering asset management services to savers. A key component of these services is wealth preservation: financial intermediaries enable investors to preserve their savings for future consumption. In doing so, financial intermediaries also enable investors to access investments that make their wealth grow over time on average. As a by-product of serving investors, intermediaries also provide investment resources to firms. We assume that investors need financial intermediaries to take advantage of these opportunities. On their own they only use highly inefficient self-storage, such as keeping money in mattresses or building houses from current income over decades without borrowing funds. Intermediaries offer savers access to financial services, such as mutual funds or mortgages, which they do not have otherwise. In Gennaioli, Shleifer, and Vishny (forthcoming; GSV), we refer to the intermediaries providing such services—be they bankers, brokers, wealth planners, or money managers—as "money doctors." The analogy captures the idea that even though generic investing in risky assets seems straightforward to economists and finance professors, it actually requires knowledge and confidence that most savers do not have. Savers rely on intermediaries to help them with financial decisions.

How do investors choose intermediaries? The central assumption of the model is that investors feel less anxious, and therefore better off, investing through intermediaries they trust. The centrality of trust in financial intermediation can be seen from financial advertising, which typically points to experience, trustworthiness, reliability, and even longevity of the intermediaries to attract investors. Guiso, Sapienza, and Zingales (2004, 2008) have pioneered empirical work showing how trust, across both investors and countries, shapes wealth allocations to risky investment. In our model, intermediaries competitively set their fees to attract clients, but because some intermediaries have a "locational" advantage of being especially trusted by some clients, in equilibrium they charge positive fees that capture a share of expected returns on investments.

GSV show that this simple model explains a range of puzzling facts about financial services. Their model explains why financial advisors are hired by investors even though they consistently underperform passive investment strategies net of fees, a major puzzle in financial economics since Jensen (1968). It explains why management fees are higher for riskier financial products that have higher expected returns. It explains why money managers pander to investor beliefs when some assets are mispriced. Here we study the aggregate implications of that model for the size of the financial sector, its costs, and their movement over time in response to shocks.

To this end, we embed a version of GSV into a Solow-style model of capital accumulation and growth under the neoclassical assumption of diminishing returns to capital. In our model, finance income tracks wealth precisely because one of its main functions is to preserve the stock of wealth, not just to finance new value added. In addition, we examine the response of the financial sector to shocks to productivity and trust. Finally, we present two extensions of the model helpful for understanding the evidence: free entry of intermediaries and population growth.

This analysis yields several principal implications. First, because the finance income share rises with the ratio of wealth to GDP, the share of finance income in GDP increases over time. The reason is that with diminishing returns to capital, there are fewer and fewer profitable projects for investing new capital along the convergence path to the steady state. As a consequence, the capital (or wealth) to GDP ratio rises, as does the finance share. Consistent with this analysis, Piketty and Zucman (2014) show that, in recent decades, the wealth to GDP ratio has increased in several advanced economies, including the U.S. We confirm that the U.S. market wealth to GDP ratio has risen over the relevant period. Piketty and Zucman (2014) explain this increase in part by slowdown in total GDP growth over time, which is precisely the mechanism we stress in our setup.

Second, our model sheds light on the evolution of costs of finance. In our model, unit fees on a given financial product fall over time because expected returns to capital fall but also because of increased competition from entry by financial intermediaries. Our model also delivers the prediction that because entry brings investors "closer" to their advisors, they take more risk over time, which might raise the unit costs of finance, since the fees on riskier investments are higher.

These predictions also find some support in the data. With respect to fees for a given financial product, French (2008) and Greenwood and Scharfstein (2013) find that management fees on equity mutual funds have fallen over time. At the same time, Philippon (2013) shows that unit costs of finance have not fallen, and Greenwood and Scharfstein (2013) document that overall fees, including those on private equity and hedge funds, have stayed roughly constant. Greenwood and Scharfstein (2013) further show that income of financial intermediaries from money management has shifted toward riskier products. These findings are consistent with our model's predictions. Indeed, we present new evidence that over time both the share of risky assets in the market portfolio and the share of U.S. population holding stocks have increased, paralleling the growth of finance share. Our model might thus help reconcile the French (2008) evidence on the declining mutual fund fees with Philippon's (2013) finding that unit costs of finance have not fallen: the reason is that investors are taking more risk at higher fees.

Third, our model ties fluctuations in the size of the financial sector to shocks in productivity and trust. In particular, our model predicts that shocks to trust immediately reduce the size of the financial sector, as investors pull resources away from their advisors. Although we do not have a model of endogenous trust determination, and hence cannot make any causal statements, some circumstantial evidence is consistent with this analysis as well. Aghion et al. (2010), Stevenson and Wolfers (2011), Sapienza and Zingales (2012), and Guiso (2010) all present evidence of sharp declines in both generalized trust and trust in the financial system during economic and financial crises. The prolonged decline of the finance share starting in the Great Depression, seen in Figure I, might be explained in part by declines in trust in the aftermath of the economic collapse.

In Section II, we describe the model. Section III presents the equilibrium in the financial sector. Section IV considers the full equilibrium in the growth model. Section V extends the model to treat endogenous entry and population growth. Section VI summarizes the empirical implications of the model and puts together some existing and some new evidence. Section VII concludes. Online Appendixes contain all proofs and also several extensions, including asset price bubbles.

II. The Model

II.A. The Household Sector

The economy is inhabited by overlapping generations of young and old. Time starts at $t = 0$ and goes on forever. A generation born at time $t-1$ contains a continuum of workers of size one, indexed by $i \in I_{t-1} \equiv [0,1]$. At $t-1$, during their young age, these workers inelastically supply their unit labor endowments at the equilibrium wage w_{t-1}. The entire wage income is saved and invested as described later, and consumption takes place only in old age after investment income is received. At the end of t, the old generation dies without bequest. We begin our analysis by considering an economy with no population growth or technical progress. This simplification allows us to focus on the money management sector, which is the novel part of our analysis.

Population growth and technical progress then affect the financial sector only indirectly, by shaping the dynamics of the per capita capital stock and the steady-state capital-to-GDP ratio. Section V.B considers population growth, and Online Appendix B.1 considers technical progress.

Workers can invest their resources in two ways. First, they can invest in self-storage. Each unit stored at $t-1$ yields $\gamma \leq 1$ units at t, so that $1-\gamma$ is lost in depreciation. We think of storage as an inefficient way to save on one's own, perhaps by holding cash or gold at home, vulnerable to the risk of theft or inflation. The case of $\gamma = 1$ captures a perfect self-storage technology. Second, a worker can hire a financial intermediary, whom we refer to as a money manager throughout, to invest his savings in a risky financial asset. At the beginning of time t, the money manager transforms a worker's savings (one for one) into capital and rents it to firms, which use it to produce output at the end of time t. We describe production in detail later.

In the model, we draw a sharp distinction between self-storage, which requires no intermediation, and risky investments, which require money managers. Self-storage can refer to keeping cash in a mattress or to building a home slowly, over years or decades, without mortgages or loans, as a form of saving (very common in developing countries). Intermediated assets are most naturally thought of as equities, but in a more general setup can include other investments. In reality, the gradation between self-storage and full financial intermediation is more continuous, from cash in mattresses to bank savings and mortgages to liquid market investments to illiquid investments such as private equity and hedge funds, with increasing amounts of intermediary attention (and cost). Our sharp differentiation is a simplifying assumption.

There are a discrete number $m > 1$ of money managers in each generation, randomly selected from the young. A generic money manager active at time t is indexed by $j \in I_t$. This money manager charges his investors a profit-maximizing fee f_{jt} per unit of investment. At time t all managers invest in the same asset, which yields a stochastic gross return R_t with mean $\mathbb{E}\{R_t\}$ and variance σ_t, both of which are determined endogenously in equilibrium. A worker/saver born at time $t-1$ delegating at the beginning of time t his risky investment to manager j thus earns a net return $R_t - f_{jt}$. If the income share invested at time t

in the risky asset is θ_t, the worker's consumption in old age is given by:

$$c_{it} = w_{t-1} \cdot \left[\gamma + \theta_t \cdot \left(R_t - \gamma - f_{jt}\right)\right].$$

Consumption increases in the excess return that risky financial assets earn relative to storage (net of the management fee). We impose the constraint $\theta_t \in (0, 1)$, which in Proposition 1 we verify to hold in equilibrium, because we are interested in cases where risk taking is interior.

After receiving the wage w_{t-1} at the end of period $t-1$, worker $i \in I_{t-1}$ chooses at the beginning of time t how much of that wage to invest in the risky asset, in storage, and which money manager $j \in \{1, \ldots, m\}$ to hire, so as to solve:

$$(1) \quad max_{j\in 1,\ldots,m,\,\theta_t\in(0,1)} w_{t-1} \cdot \left[\gamma + \theta_t \cdot \mathbb{E}\left(R_t - \gamma - f_{jt}\right) - a_{ij} \cdot \theta_t^2 \cdot \frac{\sigma_t}{2}\right].$$

The preferences of workers are mean-variance with respect to the return of their portfolio.[1] Critically, the utility of the investor i depends on the identity of manager j through the fee f_{jt} charged by j and through the manager-investor specific risk aversion parameter $a_{ij} > 1$, which we think of as the anxiety i experiences investing with j. As in GSV, saver i sees risk as being more costly with manager j, anxiety a_{ij} as higher, the lower is the trust of i for j. Investors are less anxious when taking risk with more trusted managers, perhaps because they know them or their representatives personally, or perhaps because they are persuaded by advertisement. We thus capture lower trust of i in j by a higher value of the anxiety parameter a_{ij}.[2]

II.B. Financial Intermediation

A worker's demand for the risky asset depends on his trust for different money managers and on the fees these managers charge. At each time t savers are uniformly distributed around

1. This utility function is more tractable than quadratic utility and avoids the unappealing feature of standard quadratic utility that the share of wealth invested in the risky asset decreases with wealth W. It is also more tractable than constant relative risk aversion, which requires log-normal returns and analytical approximations that complicate optimal fee setting by money managers.

2. This formulation leads to the normative conclusion that growth of finance is socially desirable even though trust creates market power distortions. An alternative, and in our view less plausible, interpretation of the model holds that delegation solely reflects investor overconfidence in the ability of managers.

the unit circle. Each manager j is also located along the circle at a constant distance $\Delta \equiv \frac{1}{m}$ from the adjacent managers. The number of managers is exogenously fixed at m (we endogenize m in Section V), and the trust of worker i in manager j is given by:

$$\frac{1}{a_{ij}} = \max(\Gamma - d_{ij}, 0), \tag{2}$$

where d_{ij} is the distance along the circle between the worker and the manager. The greater the distance between worker i and manager j, the lower is trust and the higher the worker's risk aversion.[3] Parameter $\Gamma \leq 1$ is a measure of generalized trust in the model and captures the maximal distance at which investor i is willing to delegate. If $d_{ij} \geq \Gamma$, the investor suffers infinite anxiety, namely $a_{ij} = \infty$, and so he only uses the storage technology. Two managers located at distance Δ compete for some investors as long as $\Gamma > \frac{\Delta}{2}$. An investor located halfway between these two managers is willing to take some risk with either of them. When $\Gamma < \frac{\Delta}{2}$, investors located in the middle suffer infinite anxiety from hiring either manager. These investors do not take any risk and each manager has a small, captive clientele. As we show later, whether generalized trust Γ is above or below $\frac{\Delta}{2}$ has interesting implications for the effect of competition on equilibrium fees.

At time t each manager sets his fee for the generation of savers born at $t-1$. This results in a profile $\boldsymbol{f_t} \equiv (f_{1,t}, \ldots, f_{m,t})$ of managers' fees.[4] Given this profile, each worker i chooses, based on his trust as described by equation (2), which manager to invest with and how much risky investment to undertake. The optimal policy of a worker $i \in I_{t-1}$ is summarized by a vector $[\theta^*_{ij}(\boldsymbol{f_t})]_{j=1,\ldots,m}$ that takes nonzero value only for the manager to whom the worker delegates his risky investment.[5] This vector is

3. Equation (2) describes investor trust in money managers. As a consequence, d_{ij} is 0 when the money manager invests his own money, but not when a saver takes risk on his own. In fact, savers neither trust themselves nor other savers for risky investment. For simplicity, we assume that investors have zero trust (their risk aversion is infinite) with respect to homemade or nonprofessionally managed risk taking.

4. Hsieh and Moretti (2003) show that the income of real estate agents rises when house prices rise because commissions are fixed by a trade group. In our model, fees are endogenously set by competing intermediaries.

5. In our model investors optimally invest only with their most trusted manager. Since all intermediaries manage the same asset, there is no diversification motive for hiring multiple managers. Formally, when investing with multiple

the solution of the investor's problem described in equation (1). The optimal investment policy depends on time only through the fees $\boldsymbol{f_t}$ set by managers at time t. This implies that at a fee profile $\boldsymbol{f_t}$, the profit earned by a generic money manager j from time t investment is given by:

$$\pi_{jt}(\boldsymbol{f_t}) = f_{jt} \cdot \left[\int_i \theta^*_{i,j}(\boldsymbol{f_t}) di \right] \cdot w_{t-1}. \tag{3}$$

We consider symmetric Nash equilibria in which each manager j sets the same optimal fee f^*_t identified by the condition:

$$f^*_t = argmax_{f_{jt}} \pi_{jt}(f_{jt}, f_{-jt} | f_{-jt} = f^*_t).$$

II.C. The Productive Sector

There are two inputs: labor and capital, available in aggregate supply $L_t = 1$ and K_t, respectively. We assume that capital can be converted back into consumption at no cost, but Online Appendix B.2 shows that our main results hold when we relax this assumption. Inputs at time t are owned by workers (labor is owned by the young born at time t, capital is owned by the old who are born at time $t-1$) and hired by firms in competitive markets. The production technology is risky. If an individual firm hires k_t units of capital and l_t units of labor it produces:

$$F(k_t, l_t) = \varepsilon_t \left[k_t + A \cdot k_t^{\alpha} l_t^{1-\alpha} \right]. \tag{4}$$

In equation (4), ε_t is an i.i.d shock with mean $\mathbb{E}\{\varepsilon_t\} = 1$ and variance σ. Uncertainty is realized at the end of period t when output is produced. The value of a firm consists of two components. The first is its value added $\varepsilon_t \cdot A \cdot k_t^{\alpha} l_t^{1-\alpha}$, where A captures the firm's total factor productivity. The second component is the capital stock k_t used in production, which the firm returns to investors undepreciated (up to the stochastic shock ε_t).[6]

At time t, before the shock ε_t is realized, firms hire capital and labor. Workers are hired on the spot market and are remunerated with a deterministic equilibrium wage w_t. The

managers, the anxiety of investor i is equal to $\sum_j a_{ij} z_j$, where z_j is the share of the overall risky portfolio θ_i invested with manager j.

6. Our results would change very little if the capital depreciation/appreciation shock was either different from the shock affecting value added or absent altogether. We can also allow for depreciation δ of physical capital, as long as such depreciation is smaller than the waste from inefficient storage, $\delta < 1 - \gamma$.

remuneration of capital is risky because it fully adjusts to the realization of the shock ε_t, and is paid to the holders of the firm's financial claims. These claims are bought by savers via money managers and pay an equilibrium return R_t with expected value $\mathbb{E}\{R_t\}$ and risk σ_t. The return R_t is competitively determined as a function of investment and the shock ε_t.

III. Equilibrium in the Money Management Sector

To solve a worker's portfolio problem and a manager's profit maximization problem, we take wages and expected asset returns as given. These variables are computed in the next section. At time t, each saver—after collecting his period $t-1$ wages—optimally chooses a money manager and an amount of risky investment to solve equation (1). If worker i selects money manager j, he invests in the risky asset a share $\theta_{ij}(f_{jt})$ of his wealth w_{t-1}. This share is given by:

$$\theta_{ij}(f_{jt}) = \frac{\mathbb{E}\big(R_t - \gamma - f_{jt}\big)}{a_{ij}\sigma_t}, \tag{5}$$

where $\theta_{ij}(f_{jt})$ is assumed to be in $(0, 1)$ (Proposition 1 verifies that this is the case). The saver invests $\theta_{ij}(f_{jt}) \cdot w_{t-1}$ in the risky asset and $\left[1 - \theta_{ij}(f_{jt})\right] \cdot w_{t-1}$ in storage. Risk taking increases in the excess return paid by the risky asset and in investor trust, but decreases in the risk σ_t of the financial asset. Consider now a worker's decision of which money manager to hire.

Figure II depicts the case with three managers, in which an investor i^* is located between managers j_1 and j_2. Consider the case when investors do not suffer infinite anxiety with either of the two closest managers, i.e., $\Gamma > \frac{\Delta}{2}$.

In this situation (and focusing on small deviations from a symmetric equilibrium), the investor chooses between the two closest managers j_1 and j_2. This implies that in setting his fee a generic manager, say, j_2, competes for investors on his right against j_1 and for investors on his left against j_3. To see the implications of this logic for fee setting, consider the general case in which an investor i chooses between his two closest managers j and j'. Denote the distance between investor i and his left-adjacent manager j by δ. Because the total distance between the two managers is Δ, the investor is located at distance $\Delta - \delta$ from his right-adjacent manager j'. In light of equation (2), these

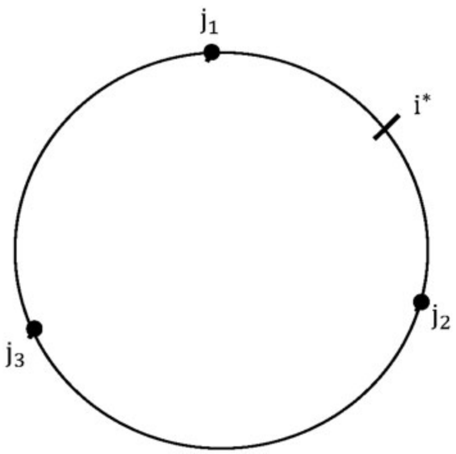

FIGURE II
Location of Investor i* with Three Managers, j_1, j_2, and j_3

distances pin down in equation (5) the investor's risky investment with either manager. By plugging these optimal risky investments into the investor's objective function of equation (1), we can show that investor i obtains a higher utility by delegating his investment to manager j rather than to manager j' if and only if:

$$\delta \leq \delta\left(f_{jt}, f_{j't}\right) \equiv \Gamma - (2\Gamma - \Delta) \cdot \frac{1}{\left[\frac{\mathbb{E}(R_t - \gamma - f_{jt})}{\mathbb{E}(R_t - \gamma - f_{j't})}\right]^2 + 1}. \tag{6}$$

Investor i thus hires manager j when the above condition holds and manager j' otherwise. Intuitively, the investor delegates his risky portfolio to manager j when his trust in j is sufficiently high, as captured by a sufficiently small distance δ from j. Other things equal, delegation to manager j is also more likely when j charges a lower fee (f_{jt} is lower) and the competing manager j' charges a higher fee ($f_{j't}$ is higher).

Consider now optimal fee setting by manager j. With the assumed circular structure, a generic manager j competes for investors against his neighbors on the left and the right. Manager j attracts investors who—according to equation (6)—are sufficiently close to him. This implies that if two competing managers

j' and j'' set the equilibrium fees $f_{j't} = f_{j''t} = f_t^*$, then the profit of manager j from setting fee f_{jt} is given by:

$$2 \cdot w_{t-1} \cdot f_{jt} \cdot \int_0^{\delta\left(f_{jt}, f_t^*\right)} (\Gamma - \delta) \cdot \frac{\mathbb{E}\left(R_t - \gamma - f_{jt}\right)}{\sigma_t} \cdot d\delta,$$

where $\delta\left(f_{jt}, f_t^*\right)$ is the maximal distance at which an investor i prefers to hire manager j at fee f_{jt} to hiring his closest competitor at the equilibrium fee f_t^*. Maximization of the above profit function yields the (sufficient) first-order condition:

$$\mathbb{E}\left(R_t - \gamma - 2f_{jt}\right) \cdot \int_0^{\delta\left(f_{jt}, f_t^*\right)} (\Gamma - \delta) \cdot d\delta + \frac{\partial \delta\left(f_{jt}, f_t^*\right)}{\partial f_{jt}} \\ \times \left[\Gamma - \delta\left(f_{jt}, f_t^*\right)\right] \cdot f_{jt} \cdot \mathbb{E}\left(R_t - \gamma - f_{jt}\right) = 0.$$

At a symmetric equilibrium $f_{jt} = f_t^*$, we obtain the following result (all proofs are in Online Appendix A).

LEMMA 1. Suppose that $\gamma > \delta/2$, which is equivalent to $m \geq \gamma/2$. Then, the equilibrium fee at time t is given by:

$$f_t^* = \left[\frac{\Delta}{\Gamma} - \left(\frac{\Delta}{2\Gamma}\right)^2\right] \cdot \frac{\mathbb{E}(R_t - \gamma)}{2} \equiv \varphi \cdot \mathbb{E}(R_t - \gamma). \tag{7}$$

where $\varphi < 1$. Management fees increase with the expected return on the risky asset. Furthermore, for $\Gamma > \frac{\Delta}{2}$, which is equivalent to $m \geq \frac{1}{2\Gamma}$, fees decrease in the number of managers m and in the generalized trust Γ that investors have in the financial sector as a whole.

From the empirical standpoint, unit fees in our model correspond to the ratio between aggregate financial sector income $f_t^* K_t$ and intermediated wealth K_t. As in GSV, equilibrium fees capture a constant fraction of the excess return expected on the risky asset. This sharing rule is intuitive: managers extract part of the surplus they enable their trusting investors to access. The fraction φ of return extracted by managers decreases as trust in all managers Γ rises. When investors trust all managers, competition among them is very intense, which drives down fees. If $m \geq \frac{1}{2\Gamma}$, fees also drop as the number of managers m rises. Intuitively, competition between highly trusted managers lowers their market power and fees. Fees fall to zero as managers fill the entire circle, namely as $m \to \infty$. In the remainder, we focus on the case where $m \geq \frac{1}{2\Gamma}$. We study the case $m < \frac{1}{2\Gamma}$ in our analysis of entry of Section V.A.

By plugging equation (7) into the optimal portfolio of equation (5), we can show that investor i places in the risky asset a share of wealth given by:

$$\theta_{ij}(f_{jt}) = \frac{(1-\varphi)\cdot \mathbb{E}(R_t - \gamma)}{a_{ij}\sigma_t}.$$

In equilibrium, each investor hires the closest manager and each manager attracts the same amount of wealth. As a consequence, the aggregate share of wealth invested in the risky asset at t, which we denote by θ_t, is the product of the number of managers m and the share of wealth managed by each of them. This aggregate share is given by:

$$\theta_t \equiv \iint_{i,j} \theta_{ij}(f_{jt}) di dj = m \cdot 2 \cdot \left[(1-\varphi)\cdot \frac{\mathbb{E}(R_t-\gamma)}{\sigma_t} \cdot \int_0^{\frac{\Delta}{2}} (\Gamma - \delta) d\delta\right] = \\ = \frac{(1-\varphi)\mathbb{E}(R_t-\gamma)}{\sigma_t} \cdot \left(\Gamma - \frac{\Delta}{4}\right), \tag{8}$$

where the expression in square brackets captures the wealth share invested by the clients to the right of a manager. With symmetry, the wealth share managed by an individual manager is twice the amount in square brackets. Equation (8) says that the share of wealth invested in the risky asset increases in the asset's excess return (net of fees) per unit of risk, in overall trust Γ, and in the number of managers $m = \frac{1}{\Delta}$. As trust in money managers increases, fees drop and investors become less anxious and are willing to take more risk.

IV. General Equilibrium Dynamics

IV.A. Production, Wages, and Asset Returns

At time t, before observing ε_t, a firm hires labor and capital to maximize expected profits:

$$max_{k_t, l_t} \mathbb{E}\left\{\varepsilon_t \cdot k_t + \varepsilon_t \cdot A \cdot k_t^{\alpha} l_t^{1-\alpha} - w_t l_t - R_t k_t\right\},$$

which are equal to total output (inclusive of both value added and the capital stock) minus factor payments. Profit maximization yields the optimality conditions:

$$(1-\alpha) k_t^{\alpha} l_t^{-\alpha} = w_t,$$

$$1 + \alpha A k_t^{\alpha-1} l_t^{1-\alpha} = \mathbb{E}\{R_t\}.$$

The marginal product of labor is equated to the wage rate, and the average marginal product of capital is equated to the average (gross) return of financial assets $\mathbb{E}\{R_t\}$.

Because the real wage is deterministic, the firm's wage bill is also deterministic, given by $w_t l_t = (1-\alpha)A k_t^{\alpha} l_t^{1-\alpha}$. The production function then implies that on the realization of a shock ε_t, the resources available to the firm's capital suppliers are $\varepsilon_t \cdot k_t + \varepsilon_t \cdot A \cdot k_t^{\alpha} l_t^{1-\alpha} - (1-\alpha)A k_t^{\alpha} l_t^{1-\alpha}$. The rate of return per unit of capital in state ε_t is therefore given by:

$$R_t = \varepsilon_t + \varepsilon_t \cdot A \cdot k_t^{\alpha-1} l_t^{1-\alpha} - (1-\alpha)A k_t^{\alpha-1} l_t^{1-\alpha}.$$

By taking the expected value of the above expression, one can immediately see that the expected return $\mathbb{E}\{R_t\}$ is equal to the average marginal product of capital $[1 + \alpha A k_t^{\alpha-1} l_t^{1-\alpha}]$, as in the first-order condition above. With constant returns to scale, remunerating capital with the residual of output after the wage bill is paid is consistent with optimality. Evaluated at the aggregate endowments K_t and $L_t = 1$, the equilibrium wage and expected return are then given by:

$$(1-\alpha)A K_t^{\alpha} = w_t, \tag{9}$$

$$1 + \alpha A K_t^{\alpha-1} = \mathbb{E}\{R_t\}. \tag{10}$$

Furthermore, by using the above expression for R_t we can show that the variance of returns is equal to $\sigma_t = var(R_t) = \sigma\left[1 + A K_t^{\alpha-1}\right]^2$.

IV.B. Evolution of the Financial Sector

We can now characterize the evolution of the economy. The total amount of risky investment at time t, which buys the aggregate capital stock K_t, is equal to the past aggregate wage bill w_{t-1} times the share of this wealth invested with money managers:

$$K_t = \theta_t \cdot w_{t-1}.$$

Using equations (9), we can rewrite this equation as:

$$K_t = \theta_t \cdot (1-\alpha)A K_{t-1}^{\alpha}. \tag{11}$$

By plugging equilibrium returns and variance into equation (8), we can compute the aggregate share of wealth invested in the risky asset, which is given by:

$$\theta_t = \frac{(1-\varphi)\left(1+\alpha A K_t^{\alpha-1}-\gamma\right)}{\sigma\left[1+AK_t^{\alpha-1}\right]^2} \cdot \left(\Gamma - \frac{\Delta}{4}\right). \tag{12}$$

Equations (11) and (12) fully characterize the dynamics of the economy. The law of motion of the capital stock in equation (11) is very similar to that obtained in a standard Solow model, with the main difference that now the amount of resources invested in the economy depends, through θ_t, on the equilibrium fees set by money managers and on the risk-return profile entailed by real investment.

In Online Appendix A we prove that, by combining equations (11) and (12) we obtain the following result:

Proposition 1. If $2\alpha > (1-\gamma)$, there are two thresholds $\bar{\sigma}$ and $\underline{\sigma}$, with $\bar{\sigma} > \underline{\sigma}$, such that, for $\sigma \in (\underline{\sigma}, \bar{\sigma})$ the economy admits a unique nonzero steady-state level of capital K^* at which individual risk taking is interior and aggregate risk taking is given by $\theta^* < 1$. The steady state is locally stable and displays the following properties:

(i) The steady-state capital stock weakly increases with the level of productivity and with the number of money managers, formally $\frac{\partial K^*}{\partial A} > 0$, $\frac{\partial K^*}{\partial m} > 0$;

(ii) Risk taking does not change with the level of productivity and increases with the number of money managers, formally $\frac{\partial K^*}{\partial A} = 0$, $\frac{\partial \theta^*}{\partial m} > 0$.

When the volatility σ of the productivity shock is intermediate, the economy monotonically converges to a unique steady state level of financial intermediation and investment.[7] The steady-state level of capital increases in productivity A. When investment becomes more productive, the wage earned by the

7. The role of production risk is intuitive. If σ is too low, people are very eager to invest in the risky asset. Some or all of them give all of their wealth to money managers, setting $\theta_{t-1}^{ij} = 1$. Condition $\sigma > \underline{\sigma}$ rules out this possibility. If σ is very high, the variance of the risky asset decreases very fast with the capital stock. This can be a source of multiplicity: some equilibria are characterized by low investment and high risk (which vindicates low investment), while other equilibria feature high investment and low risk (vindicating high investment). Condition $\sigma < \bar{\sigma}$ rules out this possibility.

young and the average return promised by money managers rise. Both effects increase financial intermediation, investment, and output in the economy. An increase in the number $m = \frac{1}{\Delta}$ of money managers also increases financial intermediation, investment, and output in the steady state. There are two reasons for this. First, when m increases, investors can find a more trusted money manager, increasing—for given fees—their propensity to invest. Second, a higher m increases competition among money managers, reducing equilibrium fees and increasing for a given level of an investor's trust the investor's risk appetite. As we show in Section V.A, higher m also increases—on the extensive margin—the number of households taking risk.

The steady state is locally stable: an economy starting below or above the steady state monotonically converges to it. Figure III graphically illustrates this convergence process.

Similarly to the standard neoclassical growth models, the main source of stability is diminishing returns to capital. As the capital stock increases, wages and national income rise. This raises the demand for financial assets by savers. The increase in financial assets further increases the capital stock and thus output next period. The growth rate of the capital stock however declines over time, because new resources are invested at progressively lower returns. Growth stops eventually and the steady state is attained.[8]

This convergence process has interesting implications for the financial sector. In particular, how do fees and money management profits change as the economy grows over time? We address these issues below.

Corollary 1. Suppose that the economy starts below the steady state, namely, $K_0 < K^*$. During the transition to the steady state:

(i) The unit fee charged by money managers, which is given by:

$$f_t^* = \varphi \cdot \mathbb{E}(R_t - \gamma) = \varphi \cdot \left[1 - \gamma + \alpha \cdot A \cdot K_t^{\alpha-1}\right], \tag{13}$$

8. Unlike in the standard Solow model, diminishing returns here are not enough to guarantee stability, because in our model risk taking by households increases as the capital stock grows. The reason is that capital deepening reduces, for any given σ, the variance of R_t. This phenomenon creates the possibility of explosive paths on which capital accumulation begets further risk taking and capital accumulation. The upper bound on the variance of shocks σ ensures stability by reducing the sensitivity of risk taking to the capital stock.

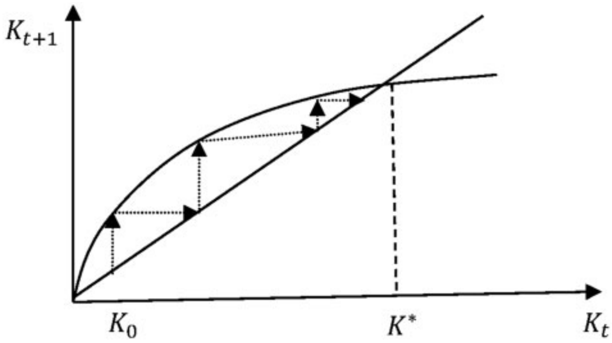

FIGURE III
Convergence to the Steady-State Capital Stock

decreases over time as capital accumulates.

(ii) The total income of the financial sector increases over time, at a higher speed than value added. The ratio of financial sector income over value added (GDP), is given by:

$$\frac{\varphi \cdot \mathbb{E}(R_t - \gamma) \cdot K_t}{AK_t^{\alpha}} = \varphi \cdot \left[\left(\frac{1-\gamma}{A}\right) \cdot K_t^{1-\alpha} + \alpha\right]. \tag{14}$$

As the economy accumulates capital, there are more resources for financial intermediation. At the same time, diminishing returns to physical capital ($\alpha < 1$) imply that *ceteris paribus* these additional resources are employed at a lower marginal return. This explains why unit management fees fall along the transition. As capital deepening reduces the expected excess return on the risky asset, it also reduces the surplus that money managers can extract from investors.

Despite this reduction in unit fees, the aggregate income earned by money managers grows over time. This is because the growth in the size of the intermediated wealth K_t more than compensates for the reduction in unit fees and causes financial sector income to rise over time. In our model financial sector income grows faster than value added, so the ratio of financial sector income to GDP grows over time. In equation (14) we exclude storage from the definition of GDP because this technology simply allows a transfer of the capital stock across periods without creating new value added in any period.[9] To illustrate our results most starkly, we also exclude from the definition of GDP the remuneration paid to finance for wealth preservation. This exclusion is also immaterial: after accounting for wealth preservation, the definition of GDP becomes $\varphi(1-\gamma)\cdot K_t + AK_t^{\alpha}$, and finance still increases as a share of GDP because $\alpha\varphi < 1$.

To understand this result, recall that in our model financial sector income can be viewed as remuneration for two services. The first is a wealth preservation service: money managers allow savers to access investment opportunities, which on average return the initial undepreciated capital and are thus better than self-storage. The second is a "growth" service: money managers enable savers to earn part of the capital income generated by these productive investment opportunities. In equilibrium, money managers are remunerated for both services. The remuneration for wealth preservation is equal to $\varphi(1-\gamma)K_t$, which is the product of the per unit of return fee φ times the surplus created by managers relative to riskless storage. Intuitively, wealth preservation is more expensive the worse is the return on riskless storage (i.e., the lower is γ). The remuneration for the growth service is equal to the per unit of return fee times capital income, namely, $\varphi\cdot\alpha\cdot AK_t^{\alpha}$. This remuneration increases in total value added AK_t^{α} and in the share α of the value added that remunerates capital. As capital stock grows, the remuneration for both wealth preservation and growth services rises, in turn increasing the aggregate income of the financial sector.

Why does the total financial income grow faster than GDP? Consider the financial sector's growth services and wealth preservation separately. As a product of real growth opportunities, income from growth services grows at the same rate as GDP. Indeed, as shown by the second term in equation (14), the remuneration for growth services is a constant fraction $\varphi\cdot\alpha$ of aggregate GDP. On the other hand, the first term in equation (14) shows that the wealth preservation service grows with the wealth to GDP ratio, and thus with the capital to GDP ratio $\frac{K_t}{Y_t}$.

9. In the text we consider the simplest case in which the elderly consume all of the current capital stock before dying. In this case, the capital stock is preserved only for one period. In Online Appendix B.2 we allow the elderly to sell their capital stock to the newborns. In this case, the capital stock is preserved for a potentially infinite period (there is no depreciation). Our predictions are not affected by the trading of the capital stock.

Finance grows relative to GDP precisely because $\frac{K_t}{Y_t}$ rises over time. In our model, this effect comes from diminishing returns: as the economy matures, the extra capital created is invested at progressively lower returns, causing the capital to GDP ratio to increase over time. The fact that a portion of the financial services is dedicated to preserving the wealth of the economy, and not to the shrinking pool of new profitable investment projects, causes the ratio of financial to total income to rise over time. This provides a novel rationale for Philippon's (2013) and Philippon and Reshef's (2013) findings that the financial sector grows relative to GDP.

Is there empirical support for our main prediction that the finance income share should grow with the wealth to GDP ratio? Figure IV presents the wealth to GDP ratio, computed for both total and financial wealth, for the United States, and shows that it rises over time. Piketty and Zucman (2014) show for several developed countries that the ratio of wealth to GDP indeed grows over long stretches of time, although they do not connect this finding to the growth of finance.

IV.C. Fluctuations in the Size of the Financial Sector

We have so far focused on long-term trends and have ignored fluctuations in the size of the financial sector, evident in Figure I. Our model also allows us to analyze the short- and long-run responses of the financial sector to shocks. We compare the effects of two permanent shocks: a permanent drop in productivity A and a drop in the overall level of trust in the financial sector Γ, owing for instance to the erosion of investor confidence during a large-scale financial crisis. Our model describes how the financial sector adjusts to these shocks.

Corollary 2. Suppose that an economy is originally in a steady state $K^*(\Gamma, A)$.

(i) Productivity A permanently drops to $A' < A$. On impact, at a given initial capital stock $K^*(\Gamma, A)$ investment drops, financial intermediation drops, but financial sector income *increases* relative to GDP. Over time, the capital stock and intermediation decrease to the new steady state $K^*(\Gamma, A') < K^*(\Gamma, A)$, and financial sector income relative to GDP returns to the initial level.

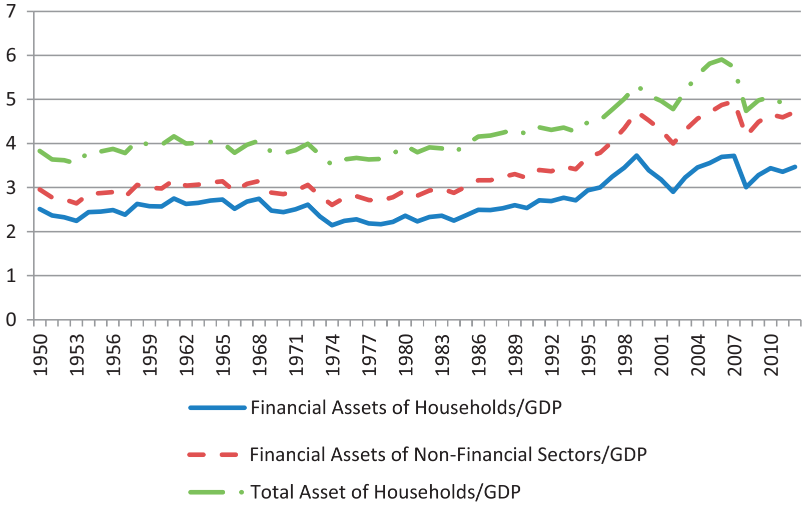

FIGURE IV

Financial Assets/GDP

Nonfinancial sectors include households, nonfarm businesses. Source of data: Philippon (2013) and Board of Governors of the Federal Reserve System (2014).

(ii) Trust Γ permanently drops to $\Gamma' < \Gamma$. On impact, at a given initial capital stock $K^*(\Gamma, A)$ investment and financial intermediation drop, and financial sector income *decreases* relative to GDP. Over time, the capital stock and intermediation gradually fall to the new steady state $K^*(\Gamma', A) < K^*(\Gamma, A)$, and financial sector income *decreases* relative to GDP.

A drop in either productivity or trust causes financial intermediation to shrink, in both the short and long run (at least weakly). In the short run, the two types of shocks entail different responses in the relative size of the financial sector. Whereas a drop in productivity causes the relative size of the financial sector to increase, a drop in trust causes the relative size of the financial sector to decline. This is because the drop in productivity reduces GDP and growth opportunities a lot but leaves the wealth preservation service of the financial sector relatively unaffected. As a consequence, the financial sector shrinks less than GDP, increasing the share of national income going to finance. In contrast, a drop in trust reduces the remuneration of both the wealth preservation and growth services of the financial sector. Although such a drop also reduces investment and income, on impact it exerts a much more dramatic effect on the financial sector income, causing the relative size of finance to drop.

In our model, permanent shocks to productivity or generalized trust can generate long-lasting boom-and-bust cycles to the size of the financial sector. As trust suddenly dissipates (owing, for instance, to a financial crisis), individuals take money out of the financial sector and put it into mattresses (self-storage). This reduces financial intermediation and the financing of profitable investment opportunities. Income reductions reduce the stock of wealth, further undermining the ability to finance investment in the future. This process generates a persistent contraction in financial intermediation and income until the new, lower equilibrium is attained.

V. Extensions

V.A. Entry into the Financial Sector

Our analysis has so far focused on the dynamics of fees and of financial intermediaries' income as shaped by the progressive exhaustion of investment opportunities (the diminishing returns assumption). In so doing, we neglected another important dimension of financial sector evolution, namely, entry of new financial intermediaries, which was precluded by the assumption that the number of money managers is fixed at $m = \frac{1}{\Delta}$.

We now allow for endogenous entry of financial intermediaries. Formally, we allow the distance Δ_t at time t between two adjacent money managers to fall over time due to entry. Denote the number of financial intermediaries at t by $m_t = \frac{1}{\Delta_t}$. For notational simplicity we treat this variable as continuous, even though the number of active managers is equal to the largest integer below m_t. We assume that creating a new money management firm at time t costs $\eta \cdot AK_t^{\alpha}$ units of consumption, where $\eta > 0$. This cost should be viewed as the value of labor that the founder must expend to set up the new financial intermediary and to earn the trust of investors (indeed, the opportunity cost of time at t is equal to the wage rate, which scales with value added).[10] Money managers can enter/exit at any time, so current

10. In this formalization, entering managers locate along the unit circle halfway between existing managers. This allows them to maximize distance from

profits are the only determinants of entry decisions. Finally, money managers appear in discrete and thus negligible numbers, so entry of additional managers leaves the labor supply of productive firms unchanged.

To investigate the effects of entry, we allow for the case in which initially some investors are so distant from money managers that they prefer not to take any portfolio risk. Denoting by $m_0 = \frac{1}{\Delta_0}$ the initial number of managers, this boils down to assuming that $\Gamma < \frac{\Delta_0}{2}$. In the simplest case where initially there are two managers ($m_0 = 2$), the distance between them is half the circle $\left(\Delta_0 = \frac{1}{2}\right)$, and $\Gamma < \frac{1}{4}$ ensures that investors located halfway between them suffer from infinite anxiety and invest everything in storage. In this case, each manager has a captive clientele and sets fees as a monopolist. As managers enter, the distance between two adjacent managers shrinks. From the first time when $\frac{\Delta_t}{2} < \Gamma$ onward, managers start competing with each other as in the case of Sections III and IV.

By generalizing our previous analysis, the proof of Lemma 2 shows that equilibrium fees at time t are now given by:

$$f_t^* = \varphi(\Delta_t) \cdot \mathbb{E}(R_t - \gamma), \text{ where } \varphi(\Delta_t) \equiv \begin{cases} \frac{1}{2} & if \quad \frac{\Delta_t}{2} > \Gamma \\ \left[\frac{\Delta_t}{\Gamma} - \left(\frac{\Delta_t}{2\Gamma}\right)^2\right] \cdot \frac{1}{2} & if \quad \frac{\Delta_t}{2} < \Gamma \end{cases}. \tag{15}$$

When there are few money managers, each of them acts as a monopolist and charges a constant fee per unit of excess return. As money managers become denser in the circle and start competing with each other, the fee $\varphi(\Delta_t)$ per unit of excess return increases in Δ_t. In this range, competition among money managers is less intense when there are fewer managers (Δ_t is higher).

If at time t a number $\frac{1}{\Delta_t}$ of money managers is active, the total profits of the financial sector are equal to $f_t^* K_t = \varphi(\Delta_t) \cdot \left[(1-\gamma)K_t + \alpha K_t^\alpha\right]$. At time t, money managers enter until the profit earned by each of them is equal to the setup cost. This condition is given by:

$$\frac{f_t^* K_t}{m_t} \equiv \varphi(\Delta_t) \cdot \Delta_t \cdot \left[(1-\gamma)K_t + \alpha A K_t^\alpha\right] = \eta \cdot A K_t^\alpha. \tag{16}$$

existing managers, which maximizes profits and allows all managers to be located at the same distance Δ_t.

By dividing both sides by AK_t^{α}, we can rewrite the equilibrium entry condition as:

$$\varphi(\Delta_t) \cdot \Delta_t \cdot \left[(1-\gamma)K_t^{1-\alpha} + \alpha\right] = \eta. \tag{17}$$

Here $\varphi(\Delta_t)$ captures the fee charged by each money manager per unit of service provided (be it wealth preservation or growth). This component increases with Δ_t because a drop in the number of managers raises fees and the aggregate income of each manager.

The second term $\Delta_t \cdot \left[(1-\gamma)K_t^{1-\alpha} + \alpha\right]$ on the left-hand side captures the share of the aggregate value of money managers' services to aggregate income provided by each individual manager at time t. As shown in the previous section, this ratio increases with the capital stock K_t because financial intermediaries' wealth preservation service becomes relatively more important as the country becomes richer. This feature drives one key property of the entry model, which we summarize in the result below.

LEMMA 2. Consider a path along which the capital stock K_t increases over time. Equation (17) implies that along this path:

(i) The number of active money managers increases (i.e., Δ_t drops) over time.
(ii) The management fees charged per unit of capital fall over time, owing both to the drop in $\varphi(\Delta_t)$ as new money managers enter, and to the fall in the marginal return to capital as K_t increases.
(iii) Entry of new managers boosts risk taking both on the extensive margin, as the number of risk taking households increases, and on the intensive margin.
(iv) The aggregate income of the financial sector increases over time, both in absolute terms and relative to the country's aggregate income.

As the capital stock expands, there are more resources available for intermediation. For given fees, money management becomes more profitable, so incurring the setup cost $\eta \cdot AK_t^{\alpha}$ becomes worthwhile. This stimulates entry of new money managers, leading to a drop in Δ_t until the profits available to an entering money manager drop back to the setup cost. In this process, managers fill the circle and increase proximity to their clients. As proximity rises, some households exclusively relying on safe storage start taking portfolio risk. In addition, competition among managers increases, driving fees down. This procompetitive effect of entry adds to the downward pressure on fees caused by capital deepening.

Despite the drop in unit fees, increased risk taking implies that the aggregate profits of the financial sector increase over time. As before, the expansion in the capital stock increases the demand for financial services. This force, which increases profits, is so strong that it more than offsets the drop in fees. Financial sector income increases not only in absolute terms but also relative to GDP. In equation (17), the left-hand side must stay constant, which implies that the total income share absorbed by finance $\varphi(\Delta_t) \cdot \left[(1-\gamma)K_t^{1-\alpha} + \alpha\right]$ increases even though higher capital stock K_t causes Δ_t to drop.

Lemma 2 considers what happens to entry and to the size of the financial sector as the capital stock K_t grows over time. We still need to verify, however, that with endogenous entry our model delivers an increasing path for the capital stock. In this case, the law of motion of the economy is still captured by equations (11) and (12) with the only difference that now also $\varphi(\Delta_t)$ and Δ_t evolve according to equation (17). In Online Appendix A we prove the following result.

PROPOSITION 2. If the parametric conditions of Proposition 1 hold, and in addition productivity A is sufficiently high, the entry model admits a unique and locally stable nonzero steady-state K^*.

Starting from initial levels of capital K_0 below the steady state, the transitional growth path is characterized by capital deepening, increasing financial intermediation, rising wealth, entry of money managers, greater participation in risky investments by households, decline in fees, and also increasing financial sector income both in absolute terms and relative to GDP. A high level of A ensures equilibrium uniqueness by bounding the role of the wealth preservation service provided by the financial sector. If A and thus the return from growth services is low, a high capital stock may alone create a strong demand for financial services, generating massive entry of intermediaries in the economy, in turn sustaining massive investment. A large A creates a sizable demand for financial intermediation regardless of the wealth preservation component, precluding the possibility of multiple equilibria.

V.B. Population Growth

We now relax the assumption of constant population (i.e., $L_t = 1$) to investigate the effect of population growth on the evolution of financial income. Suppose that the number of newborns grows at rate $n > 0$ from one generation to the next. Labor supply then satisfies the law of motion:

$$L_t = (1+n)L_{t-1}.$$

Denote by $\hat{K}_t \equiv \frac{K_t}{L_t}$ the capital stock per worker. The real wage and the expected return to capital are respectively given by:

$$(1-\alpha)A\hat{K}_t^{\alpha} = w_t,$$

$$\mathbb{E}\{R_t\} = 1 + \alpha A\hat{K}_t^{\alpha-1}.$$

with a variance of $\sigma_t = var(R_t) = \sigma[1 + A\hat{K}_t^{\alpha-1}]^2$.

To characterize the effect of n, note that the previous equations imply that the share of wealth invested in the risky asset depends on the per worker capital stock as follows:

$$\theta_t = \frac{(1-\varphi)(1+\alpha A\hat{K}_t^{\alpha-1} - \gamma)}{\sigma[1 + A\hat{K}_t^{\alpha-1}]^2} \cdot \left(\Gamma - \frac{\Delta}{4}\right).$$

The capital stock K_t employed at time t is then equal to the risky asset share θ_t times the total wage bill paid to workers at $t-1$, namely, $K_t = \theta_t \cdot w_{t-1} \cdot L_{t-1}$. Dividing both sides by L_t, and using the expression for w_{t-1}, we find that capital per worker evolves according to:

$$\hat{K}_t = \frac{\theta_t}{(1+n)} \cdot (1-\alpha)A\hat{K}_{t-1}^{\alpha}.$$

The only difference from the law of motion described in equation (11) is that now the fraction of wealth invested in the risky asset is scaled down by population growth $(1+n)$. Several immediate consequences follow. First, the capital stock per worker monotonically converges to a nonzero steady-state value $\hat{K}^*$ that is a decreasing function of n. In this steady state, output per worker and the extent of risk taking θ_t are also constant.

Second, the comparative static properties described by Proposition 1 continue to hold with respect to the steady-state levels of capital per worker and of the extent of risk taking. The transitional growth of finance income also does not change from Corollary 1. In particular, the management fee per unit of capital declines over time as $\hat{K}_t$ increases toward its steady-state level and financial sector income rises faster than value added. Critically, now the steady-state capital to GDP ratio (and thus the steady-state finance income share) increases as population growth n falls. In this sense, declining population growth also helps account for an increasing finance share.[11]

VI. Empirical Predictions

Our model yields several empirical predictions, some consistent with the available evidence, some new. The key equation of our model is:

$$\textit{finance income as a share of GDP} = f_t\theta_t\left(\frac{W_t}{Y_t}\right),$$

where W_t is aggregate wealth at t and $f_t\theta_t$ is the cost of intermediation. This equation breaks down the analysis of the dynamics of the financial sector into three components: the dynamics of the wealth to GDP ratio $\frac{W_t}{Y_t}$, the dynamics of fees f_t, and the dynamics of risk taking θ_t. Here are our main predictions concerning these components.

(1) The income share going to finance increases in the wealth to GDP ratio. The wealth to GDP ratio (which is monotonic in $\frac{K_t}{Y_t}$ in our basic model) increases as GDP growth decelerates, and decreases when some capital is destroyed (e.g., in wars).
(2) Fees f_t for a given financial product decline with the wealth to GDP ratio.

11. Population growth also allows us to analyze the role of rational asset price bubbles, which arise naturally in OLG models (Samuelson 1958; Tirole 1985). As we show in Online Appendix C, rational bubbles can help account for the evidence on the importance of valuation effects for the growth of wealth in Piketty and Zucman (2014), and may also help explain why the finance share has grown faster than our model predicts.

(3) As the wealth to GDP ratio rises, entry of new intermediaries induces households to reallocate their portfolios toward riskier, and thus more intermediated, assets (θ_t goes up). This effect may increase the average fee paid to money managers.
(4) Fluctuations in trust influence financial income through fees, wealth allocation to risky products, and the long-run level of wealth.

Prediction 1 is due to the fall in the capital income ratio during transitional growth in our neoclassical model. As economic growth slows down, the role of wealth preservation goes up, increasing the finance income share. Prediction 1 can account for the Philippon (2013) finding of the rising finance share in the United States. Piketty and Zucman (2014) show that part of the rise of $\frac{K}{Y}$ in the United States and other developed economies is precisely due to the slowdown in aggregate economic growth. According to Penn World Tables, annual U.S. per capita real GDP growth was 2.27% during 1950–1970, 2.18% during 1970–1990, and only 1.38% during 1990–2010. Over the same periods, annual population growth was 1.49%, 0.98%, and 1.07%, respectively, so *total* GDP growth has slowed down over this period from 3.86% to 3.20% to 2.51%.

For Japan and European countries, Piketty and Zucman attribute the growth slowdown primarily to declining population growth. In our model, a decline in population growth indeed renders diminishing returns to capital more severe, as shown in Section V.B. This effect increases the steady-state capital to income ratio, thereby raising the finance income share. Another prediction concerns the role of wars. Both the United States and Canada experienced declines in the finance income shares during World Wars I and II (Philippon and Reshef 2013), but neither country experienced significant war destruction. Philippon and Reshef also present some supportive data for Belgium, Spain, and the United Kingdom, but lack of data prevents a systematic analysis across countries.[12]

12. Piketty and Zucman (2014) suggest that valuation effects that are absent in our basic model explain a significant share of the increase in the U.S. wealth to GDP ratio in the past 40 years. The model of bubbles in Online Appendix C may help explain this evidence.

Prediction 2 comes from the combination of two forces: diminishing returns to capital and entry of new intermediaries. As capital accumulates, the expected return on capital declines, and equilibrium fees, which are a share of expected return, decline as well. This prediction raises the question of whether expected returns have in fact declined in the data. Although both stock and bond market returns are extremely volatile, our reading of the available research suggests that real interest rates (Campbell, Shiller, and Viceira 2009) and estimated equity premia (Campbell 2008) have declined steadily and substantially in recent decades.

A second and probably more important reason for falling fees is that, as the wealth to income ratio rises, the financial sector becomes more profitable, which induces entry of new money managers. As a result of such entry, the supply of trusted money managers increases. This effect intensifies competition among money managers in fee setting, leading to a decline in unit fees for a given product. In line with this prediction, French (2008) and Greenwood and Scharfstein (2013) document the decline in management fees over time.

Prediction 3 is due to entry of new money managers. By increasing the proximity of money managers to investors, entry increases risk taking and the size of the financial sector. On the extensive margin, entry increases the number of risk-taking households. On the intensive margin, entry increases the portfolio risk taken by each household. Increasing participation into risk taking in turn implies that despite the reduction in the equilibrium unit fee f_t^*, the unit cost of financial intermediation may actually increase as the financial sector expands. To see this, note that the total amount of financial assets in the economy at time t, which includes both storage and risky assets, is equal to w_{t-1}, the total wealth of the elderly. At the same time, the total income absorbed by the financial sector is equal to the fee times risky investment $f_t^* K_t = f_t^* \theta_t \cdot w_{t-1}$, where θ_t is the wealth share that the elderly allocate to risk taking. The unit cost of financial intermediation is then given by:

$$\frac{f_t^* \theta_t \cdot w_{t-1}}{w_{t-1}} = f_t^* \theta_t.$$

As the financial sector grows, unit fees f_t^* fall but the composition of investment shifts toward riskier assets: θ_t rises. As we show in

Online Appendix B3, the latter effect may actually dominate, causing unit costs of intermediation to rise over time.

Increased risk taking can help reconcile the French (2008) finding that in the past 30 years unit fees have come down for equity mutual funds with Philippon's (2013) evidence that the unit cost of finance have stayed roughly constant, or have even increased slightly. Some new evidence we have assembled is consistent with the increased risk taking by investors over time, which would help explain nondecreasing unit costs.[13] Figure V presents the ratio of risky assets to total financial assets in the United States since the 1950s. The figure shows a sharp rise of that ratio in the 1980s and 1990s, driven primarily by the rise in stock market valuations, but interrupted in the 2000s during the period of rapid growth of (supposedly) safe assets.

Perhaps even more dramatically, Figure VI presents the share of the U.S. population holding stocks. We have compiled this figure by pulling together various sources, including McCoy (1927), Bernheim and Schneider (1935), United States Congress (1940), Blume, Crockett, and Friend (1960), as well as the New York Stock Exchange and Federal Reserve. Figure VI shows a sharp rise in the share of population holding stocks in recent decades, paralleling the growth of finance income share. The shift toward higher risk taking thus emerges as a plausible explanation of the Philippon unit cost puzzle. Indeed, the similarity between Figures I and IV (the correlation between the two series is .87) also points to individual investing as the ultimate source of the growing finance share.

Prediction 4 holds that fluctuations in financial income are driven in part by fluctuations in trust. Coming up with causal evidence on the effects of trust on financial markets is tricky, since market fluctuations are themselves likely to affect trust,

13. We emphasize the role of stock market investment, but the argument also applies to financial innovations, which are more likely to be accepted by investors when they are close to and hence more trusting of their managers. To see the logic behind this intuition, note that in our model unit fees increase in the return paid by the financial asset (see equation (7)). Intuitively, the higher the return earned by investors, the higher the rent that the trusted money manager can extract. As a consequence, the introduction of a higher return–higher risk financial asset by a trusted money manager allows the manager to charge higher fees to investors, increasing the average cost of intermediation. We leave the formal analysis of multiple risky assets to future research.

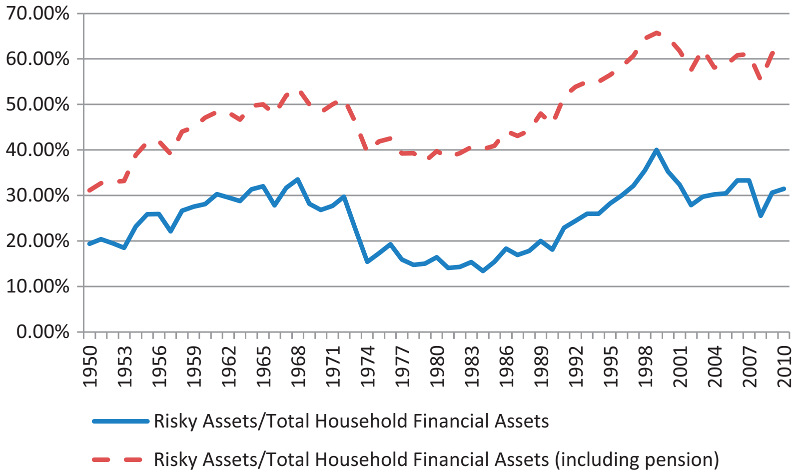

FIGURE V

Risky Asset Share in Financial Assets

Risky assets include corporate equities, mutual fund shares, syndicated loans, mortgages, student loans, and security credit held as assets by the household sector. Source: Board of Governors of the Federal Reserve System (2014).

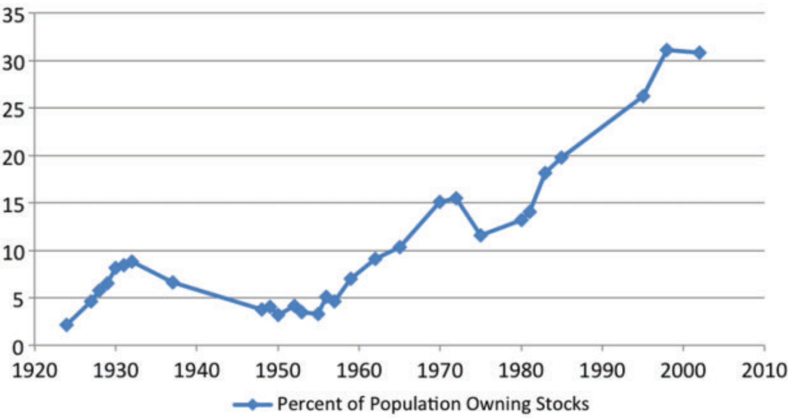

FIGURE VI

Share of Population Owning Stocks

Sources: McCoy (1927), Bernheim and Schneider (1935), United States Congress (1940), Blume et al. (1960). Board of Governors of the Federal Reserve System (2014), NYSE (2014).

so we are looking at two-way causality. Indeed, Stevenson and Wolfers (2011) present clear evidence that trust in banks in the United States declined both during the savings and loans crisis in the early 1990s and the financial crisis of 2008. But while the evidence is not definitive, trust might help shed light on some of the features of Figure I.

Specifically, Corollary 2 may help make sense of the one dramatic fluctuation in the size of the financial sector in the United States, namely, the collapse of its income from 6 to 2 percent of GDP in the Great Depression, which took 40 years to fully reverse (Figure I). The Great Depression in all likelihood combined a decline in productivity with a sharp decline in trust in the financial system. Corollary 2 suggests that both of these factors should have led to a progressive decline in the total amount of intermediated wealth. On the other hand, the fact that the income share going to the financial sector immediately shrunk underscores the role of the decline in trust. The dramatic evidence of the decline in stock market participation in the Great Depression in Figure VI is also broadly consistent with declining trust in the financial system.[14]

Malmendier and Nagel (2011) present persuasive evidence that the effects of poor market performance on investor willingness to take risk are extremely long-lasting. They interpret their findings as an effect on risk aversion, which is consistent with our idea that risk aversion is in part determined by trust. The advantages of trust as the mechanism that holds the various pieces of the model together are, first, that there exist direct measures of trust, so some of our predictions can be tested using trust data, and, second, that changes in trust have predictions for the market structure of the industry that the simple risk-aversion model does not have (GSV).

The role of trust is also consistent with the fact that the financial sector started to grow again only after World War II and reached its prewar size only in the 1980s, decades after the productivity and the wealth of the U.S. economy have substantially surpassed their pre-Depression levels. The improvements in trust over long time periods might have come from restored reputations for probity, but perhaps also from government regulation,

14. Guiso (2010) documents a decline of trust in the financial system during the S&L crisis in the United States. We have found evidence of bank deposit outflows during both this period and, on a larger scale, the Great Depression.

such as deposit insurance and securities laws.[15] The evidence of growth in stock market participation in Figure VI is consistent with this prediction as well. The slow decades-long return of trust enabled the financial sector to reach new heights as the wealth of the U.S. economy expanded.

VII. Conclusion

We have presented a Solow-style growth model in which the financial claims on the capital stock are managed by professionals. In that model, the size of the financial sector depends on both the economy's GDP and its stock of capital or wealth. The model accounts for some key facts about the development of the financial sector in the past century.

To begin, the model explains why financial sector has grown relative to GDP over time (Figure I from Philippon 2013). The reason is that one of the functions of finance is to preserve the existing stock of wealth, and wealth has grown over time relative to income, as one would expect along the adjustment path to the steady state. The model thus is also consistent with the growth of the wealth to GDP ratio over time, shown in Figure IV and more broadly by Piketty and Zucman (2014).

Our model also seeks to reconcile the somewhat conflicting evidence on the fees and unit costs of the financial sector. French (2008) presents evidence that fees on equity mutual funds have declined over time, whereas Philippon (2013) finds no evidence of declining "unit cost" of finance. According to our analysis, an important by-product of economic growth, entry by financial intermediaries, and reduction in fees is that investors allocate increasing shares of their wealth to intermediated financial products, rather than to self-storage. This implies that the composition of investor portfolios shifts over time to riskier, and hence more expensive, financial products. This can lead to increases in unit costs, even as fees for given products decline. In line with this view, we have presented in Figures V and VI some direct evidence

15. It has been suggested to us that government regulation during the Great Depression can explain the reduction in the size of the financial sector. However, the preponderance of evidence from the United States and the rest of the world shows clearly that financial regulation such as securities laws and deposit insurance is associated with stronger rather than weaker financial development (e.g., La Porta et al. 1998; La Porta, Lopez-de-Silanes, and Shleifer 2006).

of increased risk-taking by households as well as of growing stock market participation.

Our model's emphasis on trust may also help explain aspects of the volatility of the financial sector. Our approach links the sharp decline of finance in the Great Depression, and its slow recovery over the following 50 years, to the rapid decline and subsequent slow recovery of trust. Part of that recovery is exogenous, as the memory of the Great Depression recedes, but part of it is also endogenous in our model, since increases in wealth encourage entry by financial intermediaries, which creates high trust relationships.

Some see the growth of finance as an indication of problems with the market economy and the financial system. Without denying the importance of rent-seeking, agency, and other problems, our article presents a more benign view. Finance *should* grow as an economy matures, because the preservation of wealth is an increasingly important function of the financial system.

Universita' Bocconi and Igier
Harvard University
University of Chicago

Supplementary Material

An Online Appendix for this article can be found at QJE online (qje.oxfordjournals.org).

References

Aghion, Philippe, Yann Algan, Pierre Cahuc, and Andrei Shleifer, "Regulation and Distrust," *Quarterly Journal of Economics*, 125 (2010), 1015–1049.

Baumol, William, "Macroeconomics of Unbalanced Growth: The Anatomy of the Urban Crisis," *American Economic Review*, 57 (1967), 415–426.

Bernheim, Alfred, and Margaret Schneider, eds. 1935. *The Security Markets* (New York: Twentieth Century Fund, 1935).

Blume, Marshall, Jean Crockett, and Irwin Friend, "Shareownership: Characteristics and Trends," Working Paper, University of Pennsylvania, 1960.

Campbell, John Y., "Viewpoint: Estimating the Equity Premium," *The Canadian Journal of Economics*, 41 (2008), 1–21.

Campbell, John Y., Robert J. Shiller, and Luis M. Viceira, "Understanding Inflation-Indexed Bond Markets," *Brookings Papers on Economic Activity* (Spring, 2009), 79–120.

French, Kenneth, "The Cost of Active Investing," *Journal of Finance*, 63 (2008), 1537–1573.

Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny, "Money Doctors," *Journal of Finance,* forthcoming (2015).

Greenwood, Robin, and David Scharfstein, "The Growth of Finance," *Journal of Economic Perspectives*, 27, no. 2 (2013), 3–28.
Guiso, Luigi, "A Trust-Driven Financial Crisis. Implications for the Future of the Financial Markets," European University Institute Working Paper ECO 2010/07, 2010.
Guiso, Luigi, Paola Sapienza, and Luigi Zingales, "The Role of Social Capital in Financial Development," *American Economic Review*, 94, no. 3 (2004), 526–556.
———, "Trusting the Stock Market," *Journal of Finance*, 63 (2008), 2557–2600.
Hsieh, Chang-Tai, and Enrico Moretti, "Can Free Entry Be Inefficient? Fixed Commissions and Social Waste in the Real Estate Industry," *Journal of Political Economy*, 111 (2003), 1076–1122.
Jensen, Michael, "The Performance of Mutual Funds in the Period 1945–1964," *Journal of Business*, 23 (1968), 389–416.
La Porta, Rafael, Florencio Lopez-de-Silanes, and Andrei Shleifer, "What Works in Securities Laws?," *Journal of Finance*, 61 (2006), 1–32.
La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny, "Law and Finance," *Journal of Political Economy*, 106 (1998), 1113–1155.
Malmendier, Ulrike, and Stefan Nagel, "Depression Babies: Do Macroeconomic Experiences Affect Risk Taking," *Quarterly Journal of Economics*, 126 (2011), 373–416.
McCoy, Joseph, "The U.S. Legion of Capitalists," *American Bankers Association Journal*, 19, no. 8 (1927), 559–560, 626–628.
NYSE (2014). Highlights of NYSE shareowner census reports. http://www.nyxdata.com/Data-Products/Facts-and-Figures.
Philippon, Thomas, "Has the U.S. Finance Industry Become Less Efficient? On the Theory and Measurement of Financial Intermediation," Working Paper, 2013.
Philippon, Thomas, and Ariell Reshef, "Wages and Human Capital in the U.S. Financial Industry: 1909–2006," *Quarterly Journal of Economics*, 127 (2012), 1551–1609.
———, "An International Look at the Growth of Modern Finance," *Journal of Economic Perspectives*, 27, no. 2 (2013), 73–96.
Piketty, Thomas, and Gabriel Zucman, "Capital Is Back: Wealth-Income Ratios in Rich Countries 1870–2010," *Quarterly Journal of Economics*, 129, no. 3 (2014).
Samuelson, Paul, "An Exact Consumption-Loan Model of Interest with and without the Social Contrivance of Money," *Journal of Political Economy*, 66, no. 6 (1958), 467–482.
Sapienza, Paola, and Luigi Zingales, "A Trust Crisis," *International Review of Finance*, 12, no. 2 (2012), 123–131.
Stevenson, Betsy, and Justin Wolfers, "Trust in Public Institutions over the Business Cycle," *American Economic Review Papers and Proceedings*, 101, no. 3 (2011), 281–287.
Tirole, Jean, "Asset Bubbles and Overlapping Generations," *Econometrica*, 536 (1985), 1499–1528.
United States. Board of Governors of the Federal Reserve System. *Federal Reserve Statistical Release*, "Z.1 Financial Accounts of the United States." June 5, 2014.