Journal Pre-proof

The Allocation of Talent: Finance versus Entrepreneurship

Kirill Shakhnov

PII: S1094-2025(21)00061-2
DOI: https://doi.org/10.1016/j.red.2021.08.002
Reference: YREDY 1104

To appear in: *Review of Economic Dynamics*

Received date: 30 January 2018
Revised date: 12 July 2021

Please cite this article as: K. Shakhnov, The Allocation of Talent: Finance versus Entrepreneurship, *Review of Economic Dynamics*, doi: https://doi.org/10.1016/j.red.2021.08.002.

The Allocation of Talent: Finance versus Entrepreneurship

KIRILL SHAKHNOV*

September 2, 2021

ABSTRACT

The rapid growth of US financial services coupled with rapid increases in wealth inequality have been focusing policy debate as to the function of the financial sector and on its social desirability as a whole. I propose a heterogeneous agent model with asymmetric information and matching frictions that produces a tradeoff between finance and entrepreneurship. By becoming bankers, talented agents efficiently match investors with entrepreneurs, but extract excessive informational rents due to contract incompleteness. Thus the financial sector is inefficiently large in equilibrium, and this inefficiency increases with wealth inequality. The estimated model with time variation in the banker capacity accounts for the simultaneous growth of wealth inequality and the financial sector in the US. The endogenous feedback between inequality and the size of the financial sector is quantitatively important.

Keywords: talent, financial sector, matching, wealth inequality

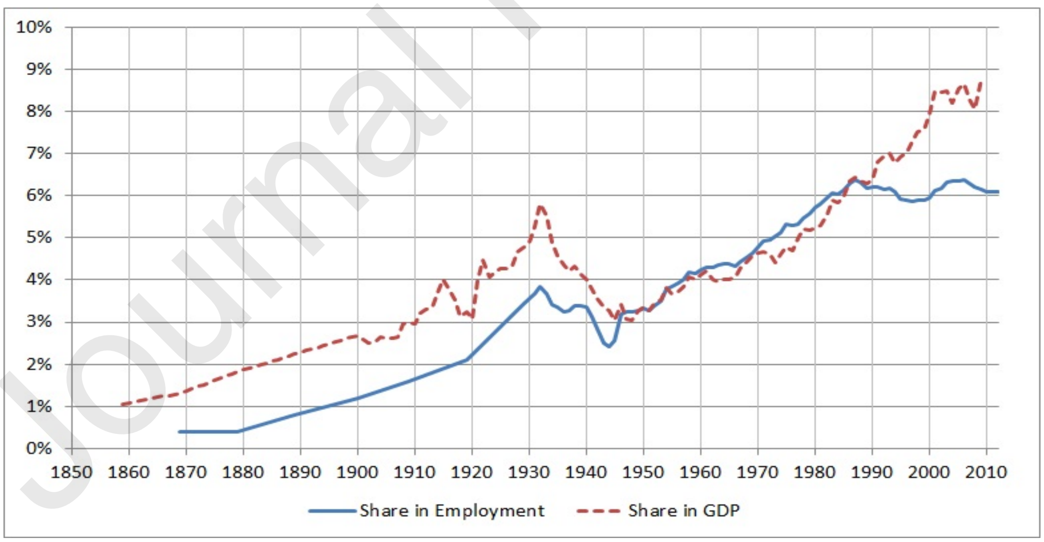
JEL classification: E44, E24, G14, L26.

*University of Surrey. I would like to thank Mark Aguiar, Juan Dolado, Boyan Jovanovic, Tim Kehoe, Omar Licandro, Rodolfo Manuelli, Claudio Michelacci, Evi Pappa and Franck Portier for many useful comments and suggestions. I am deeply indebted to Árpád Ábrahám and Ramon Marimon for all their advice and guidance. The paper has benefited from discussions with seminar participants at the University of Bonn, Bilkent University Ankara, University of Kent, New Economic School Moscow; VU University Amsterdam, Collegio Carlo Alberto, Cardiff University, EIEF Rome, EUI Florence, University of Konstanz; the Annual Congress of the European Economic Association at the University of Mannheim, UniCredit & Universities Young Economists Conference in Belgrade and the Workshop on Dynamic Macroeconomics

Introduction

> "We are throwing more and more of our resources, including the cream of our youth, into financial activities remote from the production of goods and services, into activities that generate high private rewards disproportionate to their social productivity."
>
> — James Tobin (1984)

The growth of the financial sector is well known and well documented. Figure 1 shows that the financial sector's share as a percentage of GDP as well as of employment has increased substantially since the Second World War. The figure shows that finance accounts for a higher share of GDP than of employment before the Second World War and after the 1980s (Philippon and Reshef, 2012). Interestingly, while the share of finance in employment has stabilized since the 1980s, its share of GDP has continued to rise. At the same time, this rise has been accompanied by substantial, and well documented, changes in wealth distribution.



Note: GDP shares are from Philippon (2015) and Employment shares are Buera and Kaboski (2012).

Figure 1. The growth of the financial sector in the US

This paper explains the growth of the financial sector by linking it to the dynamics of wealth distribution through an occupational choice. While the literature on occupational choices and long-run wealth distribution is well established (Banerjee and Newman (1993); Galor and Zeira (1993), and many others), few focus on finance-related occupations despite their natural association with wealth. Gennaioli et al. (2014) partially attributes the growth of finance to the increase of wealth to income ratio. Their idea is that the more wealth there is, the more assets there are to intermediate. Quantitatively, however, growth of the wealth to income ratio alone explains only a small fraction of the increase in the size of finance. I focus not on aggregate capital accumulation but rather on increasing wealth inequality and show that the growth of wealth inequality significantly contributes to the growth of finance. This echoes Piketty and Zucman (2014)'s argument that one of the reasons for increased inequality is the fact that financial services associated with asset management generate superior returns and disproportionately affect the wealthy. According to Greenwood and Scharfstein (2013), much of the growth of the financial sector comes from asset management, which is mostly a service for wealthy individuals. Furthermore, over a cross-section of countries, there is a positive relationship between inequality and the size of the financial sector.

I build a model in which financial intermediation potentially enhances welfare but draws some talented individuals away from production. The model includes three key elements: (a) heterogeneous agents who differ in terms of capital and talent; (b) an occupational choice between being a banker or an entrepreneur; (c) financial frictions. Heterogeneity and occupational choice provide a framework to study the allocation of capital (wealth) and talent. Talent is essential for both industry and the financial sector: more talent in industry means that more output is produced per unit of capital, while more talent in finance means that capital is potentially allocated more efficiently. Financial frictions in the form of private information result in the misallocation of capital as investors cannot distinguish between talented and ordinary entrepreneurs. Since it is the role of talented bankers to be able to make this distinction, the financial sector should serve to correct this misallocation.

The model generates two important insights into the financial sector. First, the model provides a novel explanation for the growth of finance by linking it to an increase in *wealth inequality*. Talented bankers provide an investment opportunity with superior returns because of their informational advantage. Only wealthy individuals can afford to pay for the services of talented bankers. In the dynamic framework, this effect is self-reinforcing: small initial differences in wealth among investors cause substantial income inequality among entrepreneurs, which is translated into greater wealth inequality during the next period. Wealthy investors are willing to pay a higher premium for financial services that increase the return on their savings, and so the greater the dispersion of wealth, the higher the price of financial services. This higher price induces a larger fraction of talented agents to pursue careers in finance. Hence, finance, wealth inequality, and inefficiency grow simultaneously.

Second, I show that decentralized equilibrium exhibits a misallocation of talent: the financial sector absorbs talent beyond a socially desirable level. Bankers create net surplus through intermediated matches and extract a part of the surplus. The size of the net surplus is proportional to the degree of wealth and talent inequality. This surplus should be split between three parties: an investor, who provides capital, an entrepreneur, who provides an idea and a banker, who efficiently matches the idea and capital. Due to matching and bargaining, there is no ex-ante reason why the split should make the income of a banker and an entrepreneur equivalent when the number of bankers is efficiently constrained. Bankers extract excessive informational rent from investors Even though the equilibrium is generically *inefficient*, efficiency can be restored by taxing the financial sector.

This second insight helps to reconcile the two sides of a debate as to whether this expansion is socially desirable. On the one hand, the former chairman of the Federal Reserve, Alan Greenspan (2002) stated: "[M]any forms and layers of financial intermediation will be required if we are to capture the full benefit of our advances in technology and trade." This idea is related to a vast literature arguing that financial development causes economic growth, because the financial sector corrects capital misallocation and consequently enhances productivity by relaxing financial constraints. On the other hand, critics of the financial sector suggest that it might have negative implications for the allocation of talent. Another former chairman of the Federal Reserve, Paul Volcker (2010) clearly stated: "[I]f the financial sector in the United States is so important that it generates 40% of all the profits in the country... What about the effect of incentives on all our best young talent, particularly of a numerical kind, in the United States?" [1]

Many papers provide indirect empirical evidence on the misallocation of talent. Data from college graduates in the US suggests that the financial sector has become one of the most popular destinations for graduates of elite universities with high levels of raw academic talent, regardless of their major (see Goldin and Katz (2008) for Harvard graduates and Shu (2015) for MIT graduates). In addition, Kneer (2013) finds that US banking deregulation reduces labor productivity disproportionately in relatively skill-intensive industries.

This paper is related to a vast literature on misallocation, particularly to papers attributing the misallocation of capital to the financial industry. (Buera and Shin, 2013; Midrigan and Xu, 2014). Whereas most studies fo-

[1] Furthermore, this concern has been vividly expressed on both sides of the Atlantic, in particular by Lord Turner, the former chairman of the UK's Financial Services Authority, who stated in 2009 that the financial sector had increased "beyond a socially reasonable size." James Tobin (1984) and Barack Obama (2012) tend to agree. Such concerns have been supported by empirical findings. For example, Arcand et al. (2015) suggest that finance starts to harm output growth when credit to the private sector reaches 100% of GDP. Other authors, such as Lucas (1988), claim that the role of finance has been overstated, and argue that it responds passively to economic growth.

cus on the impact of frictions on output and the allocation of capital and abstract away its impact on the labor market (Jovanovic and Szentes (2013) is one of the exceptions), this paper argues that financial development has a significant impact on the allocation of both capital *and* talent, which cannot be neglected. This argument is in line with a recently growing body of theoretical literature, which studies occupational choice and rent-seeking within the financial industry (Philippon, 2010; Bolton et al., 2016). Underlying this concern is the view that finance is a largely rent-seeking industry and that the resources it attracts could be better employed elsewhere. On the empirical side, Cochrane (2013); Greenwood and Scharfstein (2013); Philippon (2015); Kurlat (2019); Bazot (2017) evaluate this argument.

Many studies analyze the causes of expansion in the financial sector. Several explanations have been suggested: fluctuation of trust in financial intermediaries (Gennaioli et al., 2014), increasing efficiency of the production sector (Bauer and Mora, 2014), and structural changes in finance itself Cooley et al. (2020). None of them provide a causal link from the increase in wealth inequality to the expansion of the financial sector. On the contrary, Greenwood and Jovanovic (1990) theoretically show that financial development might cause a reduction in inequality. Cooley et al. (2020) develop a model of human capital accumulation and increasing competition for talent in the financial sector that generates more risk-taking, greater income inequality, and higher aggregate income.

There is a vast literature that studies the relationship between capital accumulation and financial development (Levine (2005) surveys the literature). Apart from the current paper, few other papers have analyzed the efficient size of the financial sector in the context of *occupational choice models*. The financial sector is inefficient in all the literature discussed below, but authors find the sources of inefficiency in very different places.

Murphy et al. (1991) argue that the flow of talented individuals into law and finance might not be entirely desirable, because law and finance provide high private but low social returns (are pure rent-seeking activities). However, they give no reason for disparities between social and private returns. The study of Philippon (2010) acknowledges the meaningful role of the financial sector, as a monitoring device that helps to overcome the opportunistic behavior of entrepreneurs in the class of occupational choice models. The source of inefficiency in the model is the production externality in the forms of human capital within the industrial sector. The projects developed by entrepreneurs have higher social than private benefits. Therefore, they need to be subsidized relative to all other occupations: workers and bankers. Bolton et al. (2016) focus on financial innovations in the sense that the financial sector creates a new over-the-counter (OTC) market. Similar to my paper, informed dealers in the OTC market extract excessive informational rent and, consequently, the financial sector attracts too many individuals. However, contrary to my model, the size of the informational rent is independent of wealth inequality.[2]

The calibrated model provides a good qualitative replication of the US data: the increase in inequality, and the growth of the financial sector as a share of both employment and value-added. While the number of talented agents limits the size of the financial sector in terms of employment, the size of the financial sector in terms of value added grows with the surplus of all matches, which each banker intermediates. An exogenous increase in the banker capacity, which can be viewed as a driver of the financial sector's productivity, is necessary to explain a large increase in the relative productivity of the financial sector. The increase in the banker capacity generates the rise not only in the size of finance, which has more than doubled

[2]Furthermore, none of these papers seek to explain the growth of the financial sector. None of them consider the financial sector to be a financial intermediary connecting investors and entrepreneurs. Neither Murphy et al. (1991) nor Philippon (2010) allow for excessive informational rent extraction, and neither Philippon (2010) nor Bolton et al. (2016) have a role for talent in either finance or industry. Several other papers (Bond and Glode, 2014; Axelson and Bond, 2015; Glode and Lowery, 2015) look at the efficiency of occupational choices within the financial sector.

since World War II but also in inequality. The endogenous feedback from inequity to the financial sector is quantitatively important, and it accounts for 20 % of the variability in the size of finance.

The paper is structured as follows. Section 1 describes the static version of the model and policy results. Section 2 provides a dynamic extension of the model and a quantitative analysis. The last section concludes.

1 Simple static model

This section presents a static version of a matching model with asymmetric information between capital- and talent-providers to study the interaction between wealth inequality and the size of the financial sector. Key elements of the model are summarized as the following: for the sake of production there are penniless talented entrepreneurs (interpreted as "young people" in the dynamic version) who have private information about the extent to which their own talent levels can create value. They seek to be paired with talentless capital providers (interpreted as "old people" in the dynamic version) who have private information about their own levels of wealth. The production technology exhibits supermodularity, so assortative matching is optimal: match highly talented entrepreneurs with wealthy capital providers.

After introducing the environment, the paper establishes two benchmarks: the first-best allocation with full information, the constrained efficient allocation with intermediation. The first one provides the upper bound on welfare. Due to the supermodularity of production, the first-best efficient outcome (without asymmetric information) is assortative matching (i.e., the most wealthy individuals paired with the most talented). However, due to private information, the first-best efficient outcome cannot be obtained. Introducing a third party with screening capabilities, which are interpreted as bankers, helps to bring the allocation close to the first best. I further assume that talented entrepreneurs are better at both producing and screening, so only talented entrepreneurs would be selected to become bankers to improve the matching efficiency in the economy, at the opportunity cost of giving up their talent in entrepreneurial activities. By optimally choosing a fraction of talented agents to work as bankers, the social planner determines the constrained efficient allocation.

Then, I study whether the constrained efficient allocation can be decentralized and show that, because talented entrepreneurs do not internalize the social benefit/cost of being a banker or entrepreneur when making their career choices, the equilibrium outcome generally features an inefficient size of the financial sector. Furthermore, when wealth inequality or talent inequality increases, the surplus of an assortative match increases, leading to higher demand for banker's services and a larger financial sector. In the dynamic version, the effect is self-reinforced.

1.1 Environment

We consider a static two-sided one-to-one matching market. The economy consists of two types of agents: investors and entrepreneurs. In order to produce output, two inputs are required: capital and an idea. Investors have wealth but no investment opportunities of their own, while entrepreneurs have ideas but need external funding.

Agents are heterogeneously endowed with talent and wealth. (Since capital is the only asset in the economy, the terms "wealth" and "capital" are used interchangeably.) Agents with talent, entrepreneurs, can choose whether to remain entrepreneurs or to become bankers instead. In industry, talent translates into capital productivity: the more talented the entrepreneur, the more output produced from a unit of capital. In finance, talent affects bankers' ability to screen entrepreneurs. See Appendix A.7 for details.

In this section, for the sake of simplicity, I consider a *very particular distribution of wealth and talent:* there is a unit mass of agents with talent and no capital, who can be talented z^H or ordinary z^L; there is a unit mass of agents with capital and no talent, who can be capital-abundant k^H or capital-scarce k^L. The share of capital-abundant investors (talented entrepreneurs) is denoted as β^i (β^e). Hence, the mass of agents with capital is equal to the mass of agents with talent. Agents with capital and no talent are potential investors, while agents with no capital and talent can be either entrepreneurs or bankers.

I consider the simplest case *one-to-one matching*. Every investor can be matched with at most one entrepreneur. Furthermore, I assume that all short-sided agents are matched with certainty. I introduce this assumption to abstract from specifying the possible advantage of a banker in terms of matching technology. The outcome of the match is given by a function, which depends on capital k and talent z:

$$F(z, k). \tag{1}$$

I assume that the function $F(z, k)$ is *strictly supermodular*. The strict supermodularity in the discrete case is given by:

$$F(z^H, k^H) + F(z^L, k^L) > F(z^H, k^L) + F(z^L, k^H). \tag{2}$$

Condition (2) suggests that positive assortative matching maximizes the sum of match outputs when the entrepreneur's type and the investor's type are complements in the match output function. For simplicity, I introduce an additional notation $F_{JI} = F(z^J, k^I)$, where $I, J = \{H, L\}$, J stands for the entrepreneur's type and I stands for the investor's type. For example, F_{HH} is the outcome of a match between a high-type entrepreneur and a high-type investor.

The economy is subject to financial frictions caused by private information, meaning that the types of entrepreneurs are not publicly observable. When investors are looking for investment opportunities, they do not know whether the entrepreneur that they meet is talented or ordinary. In an economy with private information, but without matching, the aggregate outcome is precisely random matching, because investors optimally allocate equal shares to every entrepreneur. Matching simply ensures that all funds are not allocated to one entrepreneur. Alternatively, we can simply assume that *without financial intermediation* the investment technology in the economy is random matching.

All agents are assumed to be risk-neutral and discount the future at a zero rate, so all agents maximize their incomes. The full dynamic model presented in the next section will incorporate the same static model into a dynamic framework. The detail discussion of all assumptions is in Appendix A.8.

1.2 Full information vs asymmetric information

In this section, I define the *first best* as an optimal allocation under full information (the types of entrepreneurs and investors are public information). The only constraint is the technological constraint of one-to-one matching. Under the supermodularity assumption on the production function for matching output and observability of types, the most efficient outcome in this economy is positive assortative matching—when high-type entrepreneurs are matched with high-type investors, and low-type entrepreneurs are matched with low-type investors.

Without loss of generality, let me further assume that the economy is a talent scarce economy. In other words, the number of talented agents is less than the number of capital-abundant investors ($\beta^e \leq \beta^i$). In this case, the *first best* aggregate output equals:

$$A\left[\beta^e F_{HH} + (\beta^i - \beta^e) F_{LH} + (1 - \beta^i) F_{LL}\right] \tag{3}$$

Under full information, there are many ways to achieve this allocation in a decentralized equilibrium. In particular, if one of the agents makes a take-it-or-leave-it offer to the other, which can accept or reject, the first best allocation is achieved. In other words, it does not matter whether the agent is an entrepreneur or an investor.

Since the financial sector mitigates information frictions but does not directly contribute to production, the first best in this economy is the allocation in which nobody is a banker, and all talented agents are matched with capital-abundant investors. The model without finance and full information is a variant of the standard static model of two-sided matching in which a Becker–Brock type of assignment problem arises. See the Becker–Brock efficient matching theorem (Becker, 1973). Because of private information about types, the assignment should be random – *without financial intermediation.*

Figure 2 shows the outcome of matches in this economy. Since investors and entrepreneurs can be of only two types, we have four possible outcomes, two of which are the same: green crosshatch lines – $F_{HH} = F(z^H, k^H)$, brown hatched oblique lines – $(F(z^H, k^L) = F(z^L, k^H))$, red dotted area – $F_{LL} = F(z^L, k^L)$.

1.3 The role of finance in the model

The role of finance in the model is to segment the capital market into two submarkets: that where all talented entrepreneurs are matched with capital-abundant investors by bankers; and that where the rest are matched randomly. In order to do so, bankers need to have superior information: they can distinguish between the types of entrepreneurs. This rests on an old idea that the financial sector has/produces superior information in comparison to ordinary investors.

Bankers are good at screening and sorting entrepreneurs, but they do not directly produce any output. The ability to screen depends on talent. Both finance and industry require talent. While talent in industry increases a firm's productivity, talent in finance gives an *advantage in obtaining information* and therefore improves sorting. This means that the financial

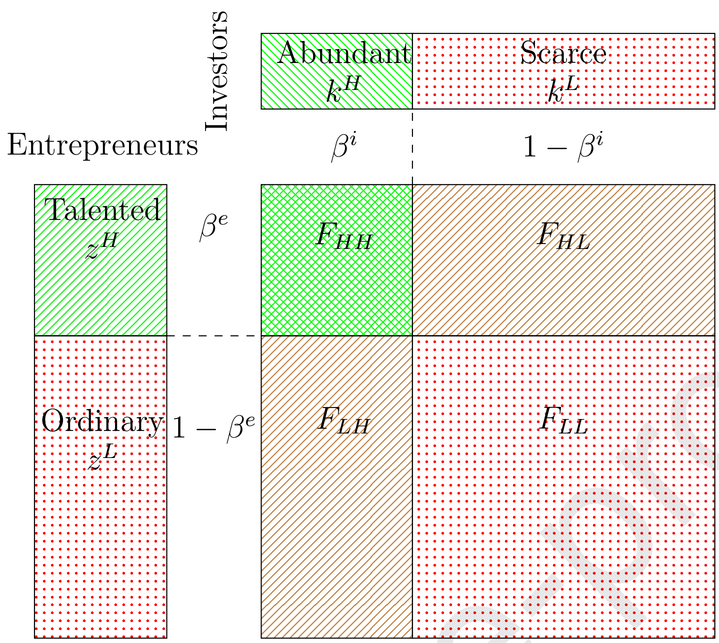


Figure 2. Model without bankers

sector brings allocation closer, but never achieves the first-best allocation. I call the allocation with financial intermediation allocation *intermediated matching*. It is important to distinguish between constrained efficient allocation under intermediated matching, discussed in the next subsection, and decentralized allocation under intermediated matching discussed later.

I consider an extreme case in which only the high-type z^H banker can be assured to match with a talented entrepreneur and a capital-abundant investor, while the low-type z^L banker can only match randomly.[3] I further

[3] This assumption accords with Philippon and Reshef's (2012) empirical observation that working in a world of innovative finance requires talent and has two possible interpretations. Under the first interpretation, the quality of sorting depends on the talent of the agent who does the sorting. A banker with ability z can distinguish between ideas with productivity z' and z'' as long as $z'' < z' \leq z$. Hence, the planner would only consider allocating talented z^H agents to finance. Under the second interpretation, there is a cost of screening $\psi(z)$ for each project discovered, which depends on talent. If this cost is high enough for the low type while low enough for the high type, $\psi(z^L) \gg \psi(z^H)$, then the planner might find it optimal to allocate to intermediation some of the talented agents,

introduce an additional technical assumption: limited capacity. A banker has neither capacity advantage nor better matching technology in comparison with ordinary investors: each banker can only provide transaction support for one deal at a time. This assumption is to ensure that one banker cannot undo all private information frictions. The only advantage that a banker has is the information advantage. This is derived by abstracting from the possible capacity advantage of intermediaries. Following Watanabe (2010), many-to-one matching can be easily introduced into the environment to accommodate the idea that another role of the financial sector is to pool risk. I relax this assumption when the dynamic setup is introduced in section 2.

To sum up, the two assumptions imply that if the share γ of talented agents β^e is allocated to the financial sector, they can match at most $\gamma\beta^e$ talented entrepreneurs. To be precise, $\min\{\gamma, 1-\gamma\}\beta^e$ talented entrepreneurs are matched by bankers with capital-abundant investors and $\max\{1-2\gamma, 0\}\beta^e$ are left for random matching. For example, in the case of $\gamma \leq 1/2$, out of talented agents β^e, the share γ is allocated to the financial sector, while the share $1-\gamma$, together with all ordinary agents $1-\beta^e$, is allocated to industry. We observe losses (the white area) because some investors remain unmatched, and gains (the green area) because the number of efficient matches increases. For simplicity, I assume that the number of capital-abundant investors is always greater than the number of bankers.Hence, some of the investors are matched with nobody.

1.4 Constrained efficiency

In this subsection, I introduce the notion of constrained efficiency. First, a social planner faces the same private information constraints as individuals do. To overcome these constraints, the planner can choose consumption of agents based on observables (the number of bankers and the outcomes of matches)

who can provide efficient matches at a small cost, while the planer would not allocate any of the ordinary agents to intermediation because of their higher matching costs.

to make sure that a fraction of talented agents selfselect themselves into the financial sector. Since only talented agents z^H can distinguish between good and bad projects, they are the only agents that need to be considered as possible bankers. By allocating the fraction γ of talented agents to finance, the planner gains the value of intermediated matches between talented entrepreneurs and capital-abundant investors F_{HH} and incurs two costs: the direct cost is due to the fact that $\gamma\beta^e$ investors become unmatched; the indirect one is that the probability of being randomly matched with talented entrepreneurs drops substantially. Second, the value of the aggregate state is irrelevant for the social planner's decision to allocate agents between occupations. For simplicity, I assume that the expected value of the aggregate state for the social planner equals μ_A. Third, because of risk neutrality the constrained efficient allocation is the one that maximizes aggregate output. The precise expression for aggregate output is given by

$$\mathbf{Y} = \mu_A \max_{\gamma} \Big\{ \tfrac{(\beta^i - \min\{\gamma, 1-\gamma\}\beta^e)}{(1-\min\{\gamma,1-\gamma\}\beta^e)} \left[\max\{1-2\gamma, 0\}\beta^e F_{HH} + (1-\beta^e) F_{LH}\right] + \tfrac{(1-\beta^i)}{(1-\min\{\gamma,1-\gamma\}\beta^e)} \left[\max\{1-2\gamma, 0\}\beta^e F_{HL} + (1-\beta^e) F_{LL}\right] + \min\{\gamma, 1-\gamma\}\beta^e F_{HH} \Big\} \tag{4}$$

As soon as γ exceeds $1/2$, all talented entrepreneurs are matched with capital-abundant investors. There is no gain to allocating an additional talented agent to the financial sector. Therefore, the constrained efficient allocation γ^* cannot exceed $1/2$; otherwise, we would observe pure losses in the number of talented entrepreneurs without any additional gains from matching, which cannot be efficient.

Proposition 1 describes the solution of problem (4):

PROPOSITION 1. *The constrained efficient allocation γ^* is always the corner solution of problem* (4)*; i.e. γ^* can be either* 0 *or* 1/2.

Proof. See Appendix A.1.

I calculate $\Delta\mathbf{Y}$, the difference between values of the planner's objective

(4) with $\gamma = 1/2$ and $\gamma = 0$. This difference is given by

$$\frac{\Delta \mathbf{Y}}{\mu_A} = (1/2 - \beta^i)\beta^e(F_{HH} - F_{HL}) - \frac{\beta^e}{2}F_{HL} - \frac{(1-\beta^i)(1-\beta^e)\beta^e}{2-\beta^e}(F_{LH} - F_{LL}). \tag{5}$$

After analyzing expression (5) above, we can conclude that, first, the contained efficient allocation is independent of the realization of aggregate states. Second, if $\beta^i \geq 1/2$, $\gamma^* = 0$ is the only possible solution to the planner's problem. For $\gamma^* = 1/2$ to be the solution, two conditions must be satisfied: $\beta^i < 1/2$, and F_{HH} needs to be adequately high. In other words, it is efficient to have a financial sector if two requirements are met: the probability of a random match between a talented entrepreneur and a capital-abundant investor is relatively low, but the value of this match is relatively high. I provide two potential interpretations of this result. One might be led to think that an economy's level of development affects the optimal size of its financial sector. In a developing country with weak institutions, it is hard for an investor to find the "right" entrepreneur. Hence, such countries need to develop their financial sectors to mitigate the effect of underdeveloped institutions. Thus the conclusion might be that the more developed the country is, the less likely it is to benefit from the financial sector. This conclusion seems at best to be counter-factual. Mayer-Haug et al. (2013) observe that entrepreneurial talent is more relevant in developing economies. Furthermore, empirical evidence suggests that developing countries suffer more from the misallocation of capital than developed ones. Nevertheless, the financial sector is efficient for countries with higher degrees of wealth or talent inequality. The more unequal a country is, the higher are the benefits from the presence of the financial sector. There is empirical support for the latter interpretation (See also Restuccia and Rogerson (2008) for the argument that resource misallocation shows up as low TFP, and Hsieh and Klenow (2009) for empirical evidence on misallocation in China and India.)

1.5 Decentralized equilibrium

In this subsection, I introduce decentralized equilibrium (DE) and compare it to the constrained efficient allocation. The presence of bankers segments the whole market into two submarkets: random matching and intermediated matching. In the first submarket without bankers, entrepreneurs and investors are randomly matched. In the second submarket, bankers mediate the matching of entrepreneurs and investors. The main difference between the DE and the constrained efficient allocation is the fact that agents' occupational choice depends on *private* returns in the two sectors, as opposed to *social* returns in the planner's case. The planner chooses how many talented agents to allocate to finance and, at the same time, how much consumption they should get. Given the information structure, this is a complicated task for the market to solve because the number of talented agents in finance affects the way the surplus is shared between three parties: an investor, an entrepreneur, and a banker. On the one hand, a surplus is created by agents in the industry (entrepreneurs). On the other hand, private information frictions create information rent that can be captured by agents in the financial sector (bankers).

For the rest of the paper, I assume that capital-abundant investors are in an excess supply in comparison to talented agents: β^i the number of capital-abundant (high-type) investors is greater than $\gamma\beta^e$ the number of bankers as well as $(1-\gamma)\beta^e$ the number of talented (high-type) entrepreneurs. It implies that talented entrepreneurs and bankers push capital-abundant investors to their outside options. The latter is a commonly observed feature in venture capital literature (Gompers and Lerner, 2000).

The timing of the problem is also important. The problem is a one-shot game. Figure 3 describes timing. First, after anticipating equilibrium outcomes, agents choose occupations and cannot reoptimize. The talented banker screens entrepreneurs until a talented one is found. If the banker succeeds, she signs a contract to seek exclusive representation of a talented

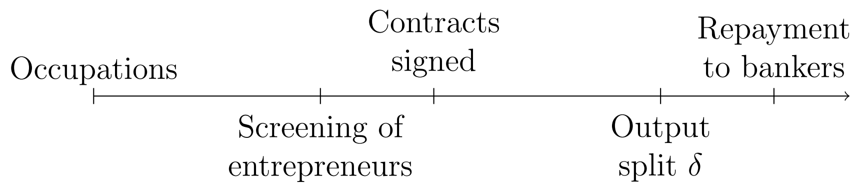


Figure 3. Timing

entrepreneur and, in exchange for fee p_e, promises to find a capital-abundant investor. Then the banker posts another contract promising for price p^i to match an investor with a talented entrepreneur. After that, all remaining investors and entrepreneurs are matched randomly. Random matching is the outside option for investors and entrepreneurs. Besides, the equilibrium of occupational choice is in pure strategies. Agents cannot play a mixed strategy to be a banker and an entrepreneur with a positive probability. After investors are matched with entrepreneurs, they bargain as to how to divide the surplus.

The most natural way to split the proceeds is *Nash bargaining*, where the bargaining power of the entrepreneur $\delta \in [0, 1]$ is exogenously given, and the bargaining power of the investor is the complement $1 - \delta$. In order to solve the Nash bargaining problem Nash (1950, 1953), one needs to define the bargaining power, the surplus of the match, and the outside options of the two counterparties. The outside option to intermediated matching is random matching. Hence, the problem must be solved *backwards*. First, I provide the solution for random matching with a given size of the financial sector γ. Then, I use the solution for random matching as outside options for the intermediated matching problem.

As a reminder, the solution of Nash bargaining contains the set of feasible utility payoffs from an agreement U and the utility payoffs to the players from a disagreement D. Since preferences are linear, the sets U and D are given by

$$
\begin{aligned}
U &= \left\{(x^e, x^i)|x^e + x^i = F(z,k), x^j \geq 0\right\}, && (6)\\
D &= \left\{(d^e, d^i)|\right\}, && (7)
\end{aligned}
$$

where x^e and x^i are the payoffs to the entrepreneur and to the investor. The entrepreneur's payoff is

$$x^e = \arg\max\left[(x-d^e)^{\delta}(F(z,k)-x-d^i)^{1-\delta}\right]. \tag{8}$$

The solutions are:

$$
\begin{aligned}
x^e &= \delta\left(F(z,k)-d^i\right)+(1-\delta)d^e, && (9)\\
x^i &= (1-\delta)\left(F(z,k)-d^e\right)+\delta d^i. && (10)
\end{aligned}
$$

As every banker can discover at most one good project, the total number of discovered good projects that are different from each other is $\min\{\gamma, 1-\gamma\}$. It is worth mentioning that, contrary to the planner's solution to problem (4), $\gamma^* \leq 1/2$, the market outcome can be any number in the interval $[0,1]$.

Assume that investors have no access to storage technology, while entrepreneurs have no opportunity for outside borrowing. Thus, the outside options for random matching — the set D in (7) — are $(0,0)$. The solution to the Nash bargaining problem gives the value of random matching for a capital-abundant investor. Note that not all investors are matched. The value of random matching is equal to the probability of matching with somebody Pr^m multiplied by the sum of products of the probability of being matched with a talented (ordinary) entrepreneur Pr^H (Pr^L) and the value of the match for a capital-abundant investor $(1-\delta)F(z^I, k^H)$. It turns out that:

$$Pr^m = \frac{1-\gamma\beta^e - \min\{\gamma, 1-\gamma\}\beta^e}{1-\min\{\gamma, 1-\gamma\}\beta^e},$$
$$Pr^H = \frac{(1-\gamma-\min\{\gamma, 1-\gamma\})\beta^e}{1-\gamma\beta^e - \min\{\gamma, 1-\gamma\}\beta^e},$$
$$Pr^L = \frac{1-\beta^e}{1-\gamma\beta^e - \min\{\gamma, 1-\gamma\}\beta^e}.$$

Hence the outside option for intermediated matching is

$$d^i = \frac{1-\delta}{1-\min\{\gamma, 1-\gamma\}\beta^e}\left[(1-\gamma-\min\{\gamma, 1-\gamma\})\beta^e F_{HH} + (1-\beta^e)F_{LH}\right] + \frac{\gamma\beta^e}{1-\gamma\beta^e}0. \tag{11}$$

Equation (11) defines the value of random matching for a capital-abundant investor, which is the outside option of a capital-abundant investor when negotiating a deal with a talented entrepreneur after intermediated matching. It is important to note that an increase in the size of the financial sector γ worsens the outside option of the capital-abundant investor because it affects the relative proportions of agents. I shall return to this point later on.

Similar to (11), the value of random matching for a talented entrepreneur, which is the outside option for bargaining in the case of intermediated matching, is

$$d^e = \frac{\delta}{1-\min\{\gamma, 1-\gamma\}\beta^e}\left[(\beta^i - \min\{\gamma, 1-\gamma\}\beta^e)F_{HH} + (1-\beta^i)F_{HL}\right]. \tag{12}$$

Applying once again the solution of Nash bargaining (9) to the intermediated matching case, I obtain a restriction on the prices that can be extracted from investors (13) and entrepreneurs (14):

$$p_i \leq (1-\delta)(F_{HH} - d^i - d^e), \tag{13}$$

$$p_e \leq \delta(F_{HH} - d^i - d^e). \tag{14}$$

Conditions (13) and (14) are the participation constraints of a capital-abundant investor and a talented entrepreneur. They state that both an investor and an entrepreneur being matched by a banker cannot be worse off in comparison to the random matching scenario. However, these inequalities are not necessarily binding, it depends on which agents are on the *short side* of the market. In addition, the prices should be non-negative.

To complete the description of equilibria, I need an additional condition (15). For the solution to be interior, $\gamma \in (0,1)$, the talented agent ($z^H > 0$) must be indifferent as to which occupation she chooses. The income of a talented banker is the probability of finding a talented entrepreneur multiplied by the sum of the two prices that are charged to an investor and an entrepreneur. As long as there are more talented entrepreneurs in the market than bankers, the probability of finding a talented entrepreneur is equal to one. The income of a talented entrepreneur is the share of the surplus received from the match with a capital-abundant investor. The indifference condition is therefore

$$Pr^b(\gamma)(p_i + p_e) = \delta \left(F_{HH} - d^i \right) + (1-\delta) d^e, \tag{15}$$

where $Pr^b(\gamma) = \frac{\min\{\gamma, 1-\gamma\}}{\gamma}$ is the probability of finding a talented entrepreneur.

Three conditions characterize all decentralized equilibria: the occupational choice condition (15) and two participation constraints in financial services, one for capital-abundant investors (13) and one for talented entrepreneurs (14). For the sake of space, I restrict attention to the case in which exogenous parameters are such that the constrained efficient size of the financial sector is *strictly positive* ($\gamma^* = 1/2$). I take the view that the financial sector is essential to the economy. Furthermore, this is an interesting case in which to study policy as policy analysis is significant in the parameter space where the financial sector plays a useful role. Proposition 2 characterizes the decentralized equilibrium in the $\gamma^* = 1/2$ case in terms of efficiency. A detailed analysis of all possible cases can be found in Appendix

A.2

PROPOSITION 2. *If it is socially efficient to have a financial sector ($\gamma^* = 1/2$) and a decentralized equilibrium exists,*

i. It is unique, $\hat{\gamma}$;
ii. This equilibrium is generically inefficient, $\hat{\gamma} \geq \gamma^*$; and
iii. There exists a restriction $\hat{\delta}$ on the set of exogenous parameters that restores the constrained efficient allocation.

Proof. See Appendix A.2.

Proposition 2 is very intuitive. Since investors are on the long side of the market, bankers extract part of the surplus from them. These profits correspond to lower levels of entrepreneurial bargaining, and the greater they are the more excessive is the entry into finance. In this case, bankers sustain equilibrium through rationing. On the contrary, for high levels of entrepreneurial bargaining power, there is no equilibrium because of the discontinuity of payoffs at $\gamma = 0.5$. If there are fewer bankers than entrepreneurs, all entrepreneurs want to become bankers because switching to the financial sector enables bankers to extract all surplus from both investors and entrepreneurs. If there are more bankers than entrepreneurs, all bankers want become entrepreneurs because it provides them with a higher income. The source of inefficiency is not the bargaining or matching friction per se, but the fact that bankers cannot cross-subsidize entrepreneurs (the price (14) should be non-negative). Potentially bankers would like to subsidize *only* talented entrepreneurs. However, it is unclear how bankers can do this because they do not know whether an entrepreneur is talented or ordinary until they meet them. On the investor's side, bankers select the capital-abundant investors by charging a strictly positive price (13) to discourage capital-scarce investors.

There exists a unique restriction on the set of exogenous parameters that restores the constrained efficient allocation. This restriction depends on both talent and wealth distributions. In other words, Proposition 2 states that the decentralized equilibrium is generically inefficient. For a given set of parameters, the solution of decentralized equilibrium is highly unlikely to be efficient.[4]

I restrict my attention to the case stated in proposition 2. Decentralized equilibrium is a function of all exogenous parameters: $\hat{\gamma} = f(\delta, \beta^e, z^H/z^L, \beta^i, k^H/k^L)$. For example, Figure 7 in Appendix A.5 presents the solution $\hat{\gamma}$ as a function of the bargaining power δ. As we can see, the decentralized equilibrium exists only for $\delta \in [0, \hat{\delta}]$; there is no solution for $\delta > \hat{\delta}$. The decentralized equilibrium coincides with the constrained efficient outcome only for one particular value of bargaining power $\hat{\delta}$.

The opposite case, in which the set of parameters is such that the constrained efficient size of the financial sector is *zero* ($\gamma^* = 0$), is discussed in Appendix A.2. The model of Murphy et al. (1991) can be viewed as a special case of the model under parameter restrictions such that $\gamma^* = 0$.

Is it possible to restore efficiency? The answer to such a question is yes. As mentioned above, the introduction of a policy instrument that directly affects one of the exogenous parameters can easily restore efficiency in the model. For example, should the planner set entrepreneurs' bargaining power to the particular value of $\hat{\delta}$, the decentralized equilibrium would become efficient. However, it is not very intuitive to think that such policies exist or would be easily introduced. See Appendix A.5.

[4] Part (iii) of Proposition 2 might look similar to the Hosios condition in the sense that the condition ensures that externalities cancel out (Hosios, 1990). In Hosios' original case, efficiency is achieved when the surplus share (bargaining power) between workers and a firm is equal to the matching share (the elasticity of the matching function). In a frictionless environment, there is a particular mechanism, directed search, that restores efficiency. However, in a frictional environment with heterogeneous agents, even directed search might not be sufficient (Shimer and Smith, 2001).

2 Dynamic model and quantitative analysis

In this section, the model is enriched along several dimensions. First, I endogenize the growth of wealth inequity through overlapping generation (OLG) structure and an increasing return to scale. As we saw above, the joint distribution of wealth and talent is an important determinant of the size of the financial sector and the degree of inefficiency. While the distribution of talent is often considered exogenous, it is difficult to think about wealth distribution as a fully exogenous object. In this subsection, I introduce endogenous wealth accumulation. The endogenous growth of wealth inequality leads to the expansion of the financial sector. The rich get richer because they can afford to pay high fees for financial services, which yield a higher return on their savings. The higher the fees, the more talented agents work in finance. Consequently, the growth of finance and the increase in wealth inequality go hand in hand. Second, allowing for the exogenous fluctuation in the productivity of the financial sector enables this simple model to match the data. Financial sector productivity is defined in terms of a banker's capacity to match investors with entrepreneurs. In the static setup, intermediated matching is one-to-one: each banker could match at most one entrepreneur to one investor. Now bankers can match N entrepreneurs to M investors. Third, I study the impact of a progressive tax system on pre-tax and post-tax inequality.

2.1 N-to-M intermediated matching

Bankers have superior capacity. Each banker can at most match N investors to M entrepreneurs. I continue to study an equilibrium under the assumption that capital-abundant investors are on the long side with respect to bankers and bankers are on the long side with respect to talented entrepreneurs. As before, γ is the fraction of talented agents who are only bankers. I denote this case as many bankers and many investors are in equilibrium: the number of entrepreneurs is less than the number of bankers: $1-\gamma < \gamma$; the total number of projects offered by bankers is less than the total number of wealthy investors: $(1-\gamma)\beta^e M < \beta^i$. After constructing the equilibrium, I verify that the assumptions hold. Under these two assumptions, the expected income of bankers is given by the expression below:

$$\sum_{k=1}^{N} Pr(k)k(Mp_i + p_e), \tag{16}$$

where P(k) is probability to discover k entrepreneurs:

$$Pr(k) = \begin{cases} \frac{1-k\gamma}{\gamma}, & \text{if } \gamma < \frac{1}{k} \\ 0, & \text{if } \gamma \geq \frac{1}{k} \end{cases}$$

Bankers look for entrepreneurs: the probability of discovering the first entrepreneur is the ratio between the number of entrepreneurs to the number of bankers $\frac{1-\gamma}{\gamma}$. The talented banker screens entrepreneurs until the banker finds a talented one. If the banker succeeds, the banker signs a contact to seek exclusive representation of a talented entrepreneur in exchange for fee p_e the Banker promises to find a capital-abundant investor. The entrepreneur is taken out from the pool of entrepreneurs. The many bankers assumption guarantees that the probability is less than one$\frac{1-\gamma}{\gamma} < 1$, and the fee is zero $p_e = 0$. The many investor assumption guarantees that Equation 13 holds as equality $p_i = (1-\delta)(F_{HH} - d^i - d^e)$. Using Faulhaber's formula, we can calculate the banker income:

$$(1-\delta)M(F_{HH} - d^i - d^e) \begin{cases} \frac{N(N+1)}{2}\left(\frac{1}{\gamma} - \frac{2N+1}{3}\right), & \text{if } \gamma \leq \frac{1}{N} \\ \frac{1+\gamma}{2\gamma^2}\frac{1-\gamma}{3\gamma}, & \text{if } \gamma > \frac{1}{N} \end{cases} \tag{17}$$

In the case $N = 1$, we obtain the familiar expression of the left-hand side of Equation 15. In the case of one-to-one matching, the equilibrium is sustained by means of rationing bankers. The realized banker income can be high if she discovers entrepreneurs or zero if she does not. Note that in the case of increased banker capacity $N > 1$, the bankers face even greater variability of income, which depends on the number of entrepreneurs they discover. Similar to the case of one-to-one matching, the social planner would like to minimize the number of bankers conditionally on all talented entrepreneurs being matched with capital-abundant investors. The constrained efficient size of the financial sector is ($\gamma^* = \frac{1}{N+1}$).

2.2 OLG structure

To introduce simple dynamics, I consider an infinite OLG model. The OLG structure seems to be natural for two reasons. First, I study the relatively long-term dynamics of inequality and the size of finance for over at least the last six decades. Second, the generation structure is well suited to the problem, because agents undergo an interesting life cycle with low-wealth when they are young age and a higher wealth in old age. The young make an occupational choice, work in one of the two sectors, and earn income. The middle-aged invest the wealth they have accumulated while young. The old consume the results of this investment.

I adopt the most basic OLG model. Every individual maximizes lifetime consumption and lives for three periods: youth, middle age, and old age. Individuals are born at time t, work at time t, receive their income at $t + 1$ and consume at $t + 2$. Individuals pass through three stages over their life cycle: working, investment, and consumption. The young are endowed with talent of z and no wealth. The young make an occupational choice either to be an entrepreneur or a banker. The middle-aged are investors because they have wealth k which they accumulated while young. The middle-aged have a choice of either being matched randomly or paying a banker price p^i in exchange for being matched with certainty to a talented entrepreneur. The middle-aged have no talent because it fully depreciated within one period.

The old consume the result of their investments.[5]

To keep two types of wealth, I consider a stand-in household that abstracts from the distinction between expected and realized income. Following Lucas and Rapping (1969) and more recent examples (Gertler and Kiyotaki, 2010), the stand-in household assumption has been a popular tool in macroeconomics to keep models tractable. I introduce the stand-in household in the following way. There is income sharing in finance. The realized income that every banker receives is the same as her expected income. Hence, all young talented agents receive the same income and become capital-abundant investors when they are old. This assumption changes nothing for expected incomes, but keeps the model tractable. If we dropped the assumption, the number of types would grow exponentially.[6] The simple model produces life-cycle behavior consistent with the data: agents with a given talent level undergo a relatively realistic life cycle with low-income working youth, high-income investment middle age, and retirement with high consumption and zero income. Individuals typically start to accumulate assets for their retirement during middle age, around the age of 40 (Gourinchas and Parker, 2002). Wealth grows rapidly over the life cycle, reaches its peak at the age of 60, and flattens out afterwards.

[5]Alternatively, due to risk-neutrality, individuals find it optimal to save their income fully and consume only in the last period. The age-related decline of cognitive abilities is a well-established fact in psychology. There is no consensus regarding the magnitude of the effect or the exact mechanism. The wealth–age profile is also well documented. Wealth grows rapidly over the life cycle and reaches its peak during one's 60s (the end of working age) and flattens or slightly declines afterward.

[6]We can think of this as an insurance scheme within the financial sector. If agents are slightly risk-averse, $u^o_{t+1} = \left(c^o_{t+1}\right)^{1-\epsilon}$, where $\epsilon \approx 0$, all bankers are willing to engage in income sharing. Notice that while the financial occupation is risky, because of banker rationing, entrepreneurship is a safe choice. All talented agents receive the same income and become capital-abundant investors in the next period because of profit sharing. Hence the share of capital-abundant investors every period, except in the first one, is equal to $\beta^i_{t+1} = \beta^e$, expression (18). The wealth of capital-abundant investors in the next period k^H_{t+1} is defined by expression (19) using the expressions for outside options in the case of intermediated matching (12) and (11) . Finally, I define the next-period wealth of capital-scarce investors k^L_{t+1}, expression (20).

For the given distribution of talent constant over time, assume an initial distribution of wealth parametrized by the share of capital-abundant investors β_0^i and their wealth k_0^H, and the wealth of capital-scarce investors k_0^L. To use the solution of the static model from the previous subsection, the evolution of wealth distribution needs to be defined. Each talented entrepreneur is matched with M_t capital-abundant investors. Owing to the stand-in household assumptions, all talented agents have the same realized income. Hence, the system of equations below defines the evolution of the wealth distribution in the model:

$$\beta_t^i = \beta^e, \tag{18}$$

$$k_{t+1}^H = M_t \delta_t \left(F_t\left(z^H, k_t^H\right) - F_t\left(z^L, k_t^H\right)\right) + M_t F_t\left(z^L, k_t^L\right), \tag{19}$$

$$k_{t+1}^L = F_t\left(z_t^L, k_t^L\right). \tag{20}$$

While there are two mechanisms behind the growth of wealth inequality, it is exacerbated by the financial sector. First, following Romer (1986), if the production function exhibits non-decreasing return to scale with respect to capital, this is very similar to the AK production function. While the increasing return on capital generates the growth of aggregate capital, the talent differentials ensure the rise of wealth dispersion. Alternatively, it is possible to generate the rise of wealth dispersion even with a decreasing return to scale production function, but one needs to assume skill-biased technological change, which disproportionally benefits talented agents. Second, an increase in the banker capacity M_t directly leads to an increase in the income of talented agents. By segmenting the markets, bankers worsen the outside options for ordinary entrepreneurs and improve those for talented ones, resulting in even greater inequality between agents. The larger inequality feeds in the size of informational rent and consequently increases the size of the financial sector.

The next subsection brings the model to the US data in an attempt to replicate the dynamics of wealth and the financial sector.

2.3 The US experience

The goal of this section is to calibrate the model and see how well it performs in explaining the evolution of finance and inequality. To discipline the model, all parameters, with exception of the banker capacity M_t, are fixed and are either directly chosen or calibrated to match several moments of interest at the beginning of the sample. The focus is on M_t, which can be viewed as a driver of the financial sector's productivity. In the baseline estimation, M_t is calibrated to match the relative productivity of finance.

The model is deterministic, has a large degree of internal persistence, and consequently is not suited to analyze the financial sector at business cycle frequency. The capital distribution is predetermined, the talent distribution is assumed to be fixed throughout all calibrations. Consequently, the total surplus of intermediated matching, as well as the income of ordinary entrepreneurs and capital-scarce investors, is predetermined. While the size of the financial sector in terms of employment is limited by the number of talented agents and the banker capacity, the size of the financial sector in terms of value added is potentially unlimited because it is proportional to the surplus of intermediated matching, which is an increasing function of the degrees of talent and wealth inequality. In other words, while the number of bankers cannot exceed the number of talented agents, bankers' income can be infinite, because it is proportional to the surplus of intermediated matching. The banker capacity M affects the banker income by changing the number of investors who pay fees for the intermediated matching.

The financial sector in the data includes Finance and Insurance, which corresponds to sector 52 according to the North American Industry Classification System (NAICS). By the definition, the labor productivity is equal to the ratio between a measure of output (gross value added) and a measure of input use (the total number of hours worked or total employment). Consequently, the relative productivity of finance is measured as the ratio between labor productivity of finance and overall labor productivity, which turns into the ratio between the shares of finance in value-added and employment.

$$\frac{A_f}{A} = \frac{Y_f}{L_f} / \frac{Y}{L} = \frac{Y_f}{Y} \frac{L}{L_f}$$

The main equilibrium condition is the occupation choice condition Equation 15. In the case of N-to-M matching, the expected income of bankers becomes a non-linear function of γ. The matrix of simulated data $\tilde{Y}(\omega)$ is generated numerically as a function of a set of parameters Ω. I select $\omega \in \Omega$ to minimize the distance between simulated data $\tilde{Y}$ and actual data Y.

$$\hat{\omega} = \underset{\omega \in \Omega}{argmin} ||\tilde{Y}(\omega) - Y||$$

In the baseline calibration, which is denoted Model 1, Y contains one time series, which is the relative productivity of finance, and three values, which are the value of the shares of finance in employment and value-added as well as a measure of inequality in 1947. The main target is the *time series* of the relative productivity of finance. The additional set of targets contains the shares of finance in employment and value-added as well as the top 10% to the average income at the beginning of the estimation period, in 1947. Post-tax income inequality, instead of wealth inequality, is used for two reasons. First, the model does not distinguish between the two. The evolution of wealth and income inequality, measured as the ratio of top 10 % average to the overall average, are fairly similar in the data as well (see Figure 9). Second, focusing on income instead of wealth inequality permits isolating the impact of taxation. In the case of Model 2, Y instead contains the entire time series of the employment share of finance, the value-added share of finance, and the measure of inequality. The two calibrations share the same restrictions imposed on the set Ω.

Table 1 reports directly chosen parameters (Panel A), estimated parameters (Panel B) along with the corresponding targets for two calibrations. Model 2 corresponds to an exercise, where I target all three time series of interest instead of only the relative finance productivity, as is the case in the baseline calibration (Model 1). Finally, Panel C reports the performance of two models concerning four time series of interest. The shares of finance in value-added and employment in 1947 correspond to the bargaining power of entrepreneur δ and the relative productivity of talented entrepreneurs $\frac{z^H}{z^L}$.

Table 1. Estimated Parameters

Parameters	*Targets*	Models	
		1	2
Panel A: Directly chosen parameters			
α^k	1	1	1
α^z	1	1	1
β^e	Top income shares in, %	0.10	0.10
z^L	Normalization	100	100
N	Investments per professionals in a typical PE firm Metrick and Yasuda (2010)	2	2
Panel B: Estimated parameters			
$\mathbb{E}[M_t]$	Relative fin. productivity	4.4	13.9
δ	Value added share in 1947	0.41	0.45
$\frac{z^H}{z^L}$	Top income in 1947	8.8	8.7
Panel C: Other moments			
R_p^2	Relative productivity	0.96	0.76
R_e^2	Employment share	0.14	0.87
R_v^2	VA share	0.70	0.98
R_i^2	Inequality	-0.14	0.38

Note: This table reports targets and the estimated parameters for two models: model (the baseline calibration and model 2. Model 2 targets jointly three times series: finance share in employment and value-added, and a measure of inequality. Panel A reports the parameters, which are directly chosen. Panel B reports the estimated parameters. Panel C reports "R squared", i.e. the proportion of the variance in the variable of interest that is explained by the two models. The variables of interest are the relative productivity of finance, the employment share of finance, the value-added share of finance, and a measure of inequity, which is the ratio of top 10% income to the average income. All parameters are estimated jointly using a version of the Simulated Method of Moments.

Apart from the time series of banker capacity in terms of investors M, which is targeted to match the relative productivity of finance, we are left with the entrepreneurial bargaining power δ and the ratio of productivities between talented and ordinary entrepreneurs $\frac{z^H}{z^L}$, which we obtain by targeting the initial values of financial value-added and inequality in 1947. The estimated value for the relative productivity of talented entrepreneurs $\frac{z^H}{z^L}$ is 4.4 in the baseline model. The value falls within the range of the estimates based on the productivity and firm size distribution. The productivity level of top 10% firms relative to the productivity of an average firm ranges from 4 (Axtell, 2001) to 9 (Bernard et al., 2003). The estimated value for the level of entrepreneurial bargaining power δ is 42%. It is hard to find precise counterfactual data for this number. Since δ determines the share of surplus in the hands of entrepreneurs, some estimates, e.g. Kaplan and Stromberg (2003), suggest that the average founders' share equals 21.3% of a portfolio company's equity value.

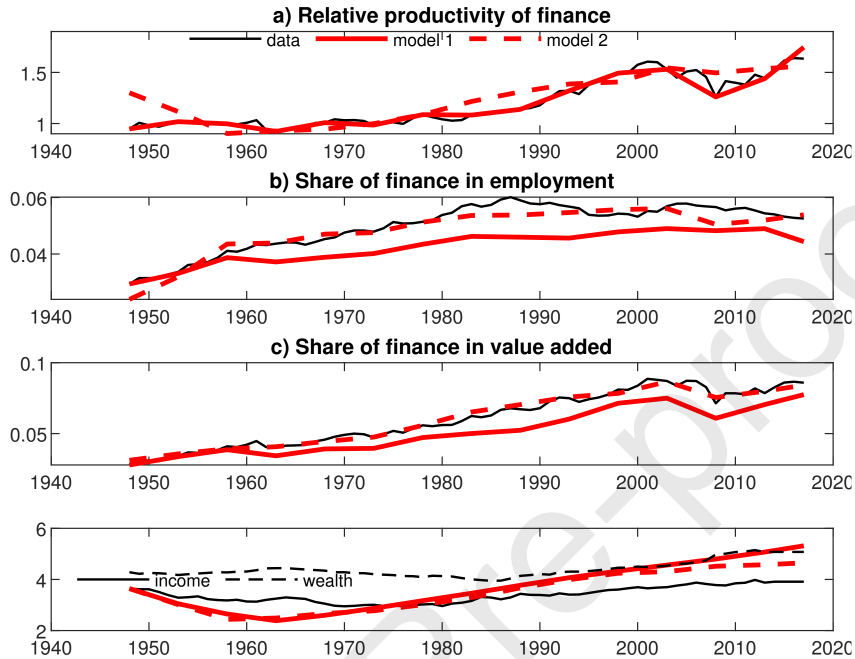
The remaining parameters are directly chosen. First, Since talented agents become capital abundant investors in the next period, the share of talented agents is set to be $\beta^i_t = \beta^e = 10\%$. Second, for both Model 1 and Model 2, I assume that the returns to scale on capital and talent are both equal to one $\alpha^k = \alpha^z = 1$, which corresponds to the production function being simply the product of two inputs $F = zk$. A particular advantage of this specification is the fact that N-to-M matching is very similar to 1-to-1 matching because an extra capital enters additively into production. Third, the bankers capacity in terms of entrepreneurs is set to be 2, which highlights the fact that a typical venture capital fund invests in a few companies, 20 on average, and employs few professional, 9 on average (Metrick and Yasuda, 2010). Hence, the resulting ratio of investments per employee is 2. Fourth, the productivity of ordinary agents z^L is normalized to be 100. In addition, the setting particular values for α^z, z^L, and N is not restrictive. What is important not the relative productivity of talented entrepreneurs per se, but the ratio taken to the power of $\left(\frac{z^H}{z^L}\right)^{\alpha_z}$. Furthermore, z^L and N do not affect the relative productivity of finance at all. See discussion of Figure 13 in appendix for more details.

Figure 4 shows the performance of Model 1 and Model 2 against targeted and untargeted moments. Even though all parameters are jointly calibrated, the rise of the relative productivity corresponds to the increase in the banker capacity M (number of investors which can be matched by a banker). The banker capacity M varies in order to match the time series of the relative productivity of finance. The main target is the time series of the relative financial productivity, which follows a well-documented pattern (Philippon and Reshef, 2012). The relative financial productivity was flat from the 1950s to 1980s, then it was rapidly increasing except for the Great Recession and reached the levels of 50 % higher than the 1980s levels by the 2000s. As seen in the figure, while the share of employment in finance was growing until the 1980s and then stabilized above 6%, the value-added share continues to grow.

This pattern of financial sector productivity is reflected in the dynamics of the calibrated banker capacity M_t. The slow growth period of the 1950s and 1970s is followed by a rapid increase during the 1980s and 2000s. Given that we can match the shares of finance in 1947, it is not surprising that the increase in the banker capacity M leads to the increase in both shares. As the banker capacity rises, their income as well as the number of bankers increases. However, the model understates the exact amount of increase. The variation in M_t explains 96% of relative productivity variation. As measured by R^2, the variation in M_t also explains 70% of the variation in the value-added share of finance and 14% of variation in the employment share of finance, both are not being directly targeted.

The inequality side, which is measured by the ratio of top 10% income to the average income, is where the performance model of baseline calibration is only qualitative. The baseline calibration matches an overall U-shape pattern

Figure 4. Targeted and untargeted series



Note: The figure compares the performance of two models against four data series. While Model 1 is the baseline calibration, where the only target is the time series of the relative productivity of finance, for Model 2, the targets are three time series, including the financial shares in employment and value-added, and a measure of inequality, which is measured as the ratio of top 10% income to the average income. The solid black line represents the data. The solid red line corresponds to model 1 and the dashed red line corresponds to Model 2. Panel a) reports the relative productivity of finance. Panel b) reports the share of finance in employment. Panel c) reports the share of finance in value-added. Panel d) plots the ratio of top 10% income to the average income. Additionally, the dashed black line represents the ratio of top 10% wealth to the average wealth in the last panel.

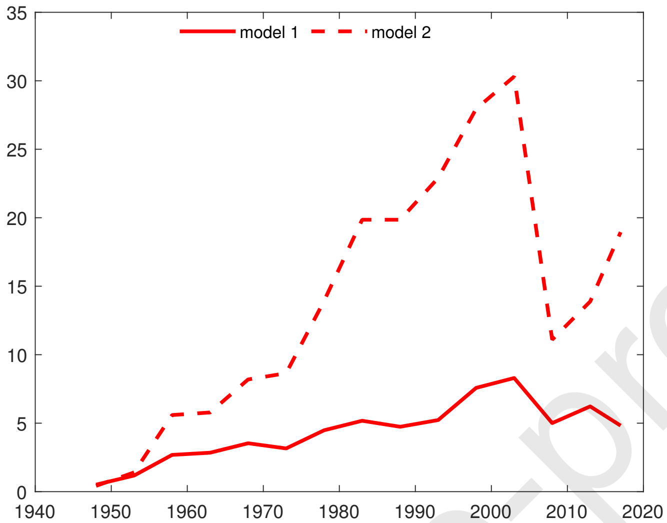
of inequality, but the movements are exacerbated. The top 10% income to the average income experiences a period of flat and slowly declining inequality during the 1950s and 1970s, which is followed by rapid expansion afterward, apart from a small drop during the Great Recession. The drop reflects the sharp decrease in asset prices: stocks, housing, etc. This business cycle point of view is beyond the scope of this paper. The inequality in the model initially falls significantly, which corresponds to the period of the stagnant relative productivity of finance, then increases rapidly and substantially contrary to the data, where only a modest initial decline is followed by a moderate subsequent increase. The fact that a large initial decline is followed by an even faster growth leads to a negative R^2 for the inequality measure.

Overall, the baseline calibration provides a good fit for the financial side of the economy explaining most of the variation in the value-added share of finance and a significant fraction of the employment share of finance. When we compare Model 1 and Model 2, it is not surprising that Model 2 outperforms Model 1 in explaining its targets, but underperforms in explaining the relative productivity of finance, which is the target for Model 1. What is surprising is how well Model 2, with only time variation in the banker capacity M_t, manages to explain the joint evolution of the financial shares in employment and value-added, and the measure of inequality. Model 2 explains 38 % of time variation in inequality and more than 90 % of financial shares. However, the performance of the model in matching both the shares of finance and the inequality can be further improved, if we add more degree of freedoms, for example, in the form of the time-varying return on capital α^k, which controls the growth of wealth

Figure 5 plots the calibrated values of M_t for the baseline model (Model 1) and Model 2 over time. In the baseline case, the banker capacity grows 20 times from the initial value to the maximum in the mid-2000s, which corresponds to 5.3 % growth annually. The increase of 71 times is even large in the case of Model 2, which corresponds to 7.6 % of the annual rise in the banker capacity. The rapid rise is followed by an even more profound drop in the case of Model 2 and only a moderate decrease in the case of Model 1. To summarize, in order to explain the joint dynamics of finance and inequality the model with only time variation in the banker capacity needs to have large swings in the banker capacity, which might be hard to justify.

In the baseline calibration, the goal is to explain the dynamics of the relative finance productivity for the whole financial sector in the post-second world war experience of the United States. This choice is driven by data avail-

Figure 5. The banker capacity



Note: This figures plot the dynamics of the banker capacity M for two calibrated models. The solid red line represents Model 1, which is the baseline calibration with the relative finance productivity is the sole target. The dashed red line represents Model 2. The targets of Model 2 are three times series: finance share in employment and value added, and a measure of inequality.

ability and the lack of large exogenous shocks during this period. For inequity data, I rely on Piketty and Zucman (2014), who calculated the median and the mean of wealth and income for top groups based on administrative data. The pre-World War II(WWII) period, in contrast to post-WWII, includes two unique events of the exogenous destruction of wealth, particularly at the top. In the United States, only the Great Depression and WWII managed to drive down wealth inequality. [7]

Then I recalibrate the model to explain the behavior of one subindustry

[7]This is similar to Wolff (2016), who calculated the median wealth and mean wealth for top groups based on the Survey of Consumer Finance and its precursor, the Survey of Financial Characteristics of Consumers. Kuhn et al. (2020) suggests that the survey data, in contrast to administrative data, provide more accurate estimates. The causes of the Great Depression and WWII are clearly outside the scope of the model and are arguable exogenously as causes of wealth inequality. Scheidel (2017) attributes this decline to the secular stagnation of housing and stock market prices relative to the price of labor.

of the financial sector—*new finance*, which includes private equity and investment banking. Even though the model can be applied to the financial sector as a whole, private equity finance is a subindustry for which the two assumptions of the model are particularly valid: matching and information superiority. In particular, it is the very business of a private equity fund to match a few selected startups with high-net-worth individuals. A private equity fund provides an opportunity to invest in a few companies with a long-term horizon for a small number of wealthy investors. Over this period new finance has grown by five times in employment and 13 times in value-added. With the fixed share of talented agents β^e in order to accommodate this rapid expansion, once again the banker capacity M needs to increase. What is more interesting is that most of the increase happens in the 1980s and 1990s, the period of a large rise in the relative productivity of the financial sector (Figure 12). See Appendix B.1 for data description and Appendix B.7 for estimation results.

The change in the banker capacity is a key determinant for the size of finance. It is robust for allowing an extra degree of freedom, or shouting down the endogenous feedback from inequality, or introducing changes to the US tax system. First, the model has a large degree of internal persistence. The capital distribution is predetermined, the talent distribution is assumed to be fixed throughout all calibrations. The banker capacity M affects the banker's income by changing the number of investors who pay fees for the intermediated matching. Another way to change the banker income is through the entrepreneurial bargaining power δ, which affects the share of the surplus, which goes to investors and eventually to bankers. The relative productivity of talented entrepreneurs $\frac{z^H}{z^L}$ affects the surplus of intermediated matching and ensures that the level of inequality is consistent with the data. All those possibilities are explored in Appendix B.3. An additional degree of freedom improves the explanatory power of the model, but do not alter the pattern of the banker capacity M_t, which is increasing over time with the acceleration around the 1980s.

Second, Appendix B.4 argues that the endogenous feedback from the size of finance on inequality and vice versa is quantitatively important. Without this feedback, the pattern of the calibrated banker capacity does not change qualitatively, but the model either overstates the growth of finance in terms of value-added or understates the growth of finance in terms of employment. This confirms the findings of the previous section, such as inequality plays an essential role in determining the equilibrium size of the financial sector Third, Changes in the US tax system and, in particular, in the reduction of tax progressivity are not quantitatively important (Appendix B.8).

3 Conclusion

This paper develops a new model of an economy with a financial sector, private information, and heterogeneous agents. The model sheds light on the role of the financial sector and its impact on the allocation of capital between entrepreneurs and the allocation of talent between finance and industry. Talent is essential for both industry and the financial sector: more talent in industry means more output is produced, while more talent in finance means capital is allocated more efficiently. The model establishes a link between the growth of the financial sector and the increase in wealth inequality. First, it shows that the market overproduces finance. Second, a sizeable increase in the banker capacity generates the growth of both the financial sector and inequality. Furthermore, the endogenous feedback from inequity to the size of finance is quantitatively important.

REFERENCES

Acemoglu, Daron, 2002, Directed technical change, *The review of economic studies* 69, 781–809.

Acharya, Viral, Marco Pagano, and Paolo Volpin, 2016, Seeking alpha: Excess risk taking and competition for managerial talent, *The Review of Financial Studies* 29, 2565–2599.

Alvaredo, Facundo, Anthony B. Atkinson, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman, 2015, The world wealth and income database, accessed online: June, 30, 2015.

Arcand, Jean Louis, Enrico Berkes, and Ugo Panizza, 2015, Too much finance?, *Journal of Economic Growth* 20, 105–148.

Axelson, Ulf, and Philip Bond, 2015, Wall street occupations, *The Journal of Finance* 70, 1949–1996.

Axtell, Robert L, 2001, Zipf distribution of us firm sizes, *science* 293, 1818–1820.

Banerjee, Abhijit V, and Andrew F Newman, 1993, Occupational Choice and the Process of Development, *Journal of Political Economy* 101, 274–298.

Bauer, Christian, and Jose V. Rodriguez Mora, 2014, The joint determination of tfp and financial sector size, Working paper, mimeo.

Bazot, Guillaume, 2017, Financial consumption and the cost of finance: Measuring financial efficiency in europe (1950–2007), *Journal of the European Economic Association* jvx008.

Becker, Gary S, 1973, A Theory of Marriage: Part I, *Journal of Political Economy* 813–46.

Bernard, Andrew B, Jonathan Eaton, J Bradford Jensen, and Samuel Kortum, 2003, Plants and productivity in international trade, *American economic review* 93, 1268–1290.

Bienz, Carsten, and Julia Hirsch, 2011, The Dynamics of Venture Capital Contracts, *Review of Finance* 16, 157–195.

Böhm, Michael J, Daniel Metzger, and Per Strömberg, 2018, 'since you're so rich, you must be really smart': Talent and the finance wage premium, *Riksbank Research Paper Series* 18–3.

Bolton, Patrick, Tano Santos, and Jose A. Scheinkman, 2016, Cream-skimming in financial markets, *The Journal of Finance* 71, 709–736.

Bond, Philip, and Vincent Glode, 2014, The labor market for bankers and regulators, *Review of Financial Studies* .

Buera, Francisco J., and Joseph P. Kaboski, 2012, The rise of the service economy, *American Economic Review* 102, 2540–69.

Buera, Francisco J., and Yongseok Shin, 2013, Financial Frictions and the Persistence of History: A Quantitative Exploration, *Journal of Political Economy* 121, 221 – 272.

Cochrane, John H., 2013, Finance: Function Matters, Not Size, *Journal of Economic Perspectives* 27, 29–50.

Cooley, Thomas, Ramon Marimon, and Vincenzo Quadrini, 2020, Commitment in organizations and the competition for talent, *The Review of Economic Studies* 87, 2165–2204.

Célérier, Claire, and Boris Vallée, 2019, Returns to talent and the finance wage premium, *Review of Financial Studies* 32, 4005–4040.

Diewert, W Erwin, and Kevin J Fox, 2008, On the estimation of returns to scale, technical progress and monopolistic markups, *Journal of Econometrics* 145, 174–193.

Durnev, Art, Randall Morck, and Bernard Yeung, 2004, Value-Enhancing Capital Budgeting and Firm-specific Stock Return Variation, *Journal of Finance* 59, 65–105.

Galor, Oded, and Joseph Zeira, 1993, Income Distribution and Macroeconomics, *Review of Economic Studies* 60, 35–52.

Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny, 2014, Finance and the Preservation of Wealth, *The Quarterly Journal of Economics* 129, 1221–1254.

Gertler, Mark, and Nobuhiro Kiyotaki, 2010, Financial Intermediation and Credit Policy in Business Cycle Analysis, in *Handbook of Monetary Economics*, chapter 11, 547–599 (Elsevier).

Gladstone, David, 1988, *Venture Capital Handbook* (Englewood Cliffs, NJ: Prentice Hall).

Glode, Vincent, and Richard Lowery, 2015, Compensating financial experts, *The Journal of Finance* .

Goldin, Claudia, and Lawrence F. Katz, 2008, Transitions: Career and family life cycles of the educational elite, *American Economic Review* 98, 363–69.

Gompers, Paul, and Josh Lerner, 2000, Money chasing deals? The impact of fund inflows on private equity valuation, *Journal of Financial Economics* 55, 281–325.

Gordon, Platt, 2000, The americas, *Global Finance* 14.

Gourinchas, Pierre-Olivier, and Jonathan A. Parker, 2002, Consumption Over the Life Cycle, *Econometrica* 70, 47–89.

Greenwood, Jeremy, and Boyan Jovanovic, 1990, Financial development, growth, and the distribution of income, *Journal of Political Economy* 98, 1076–1107.

Greenwood, Robin, and David Scharfstein, 2013, The Growth of Finance, *Journal of Economic Perspectives* 27, 3–28.

Guvenen, Fatih, Burhanettin Kuruscu, and Serdar Ozkan, 2014, Taxation of human capital and wage inequality: A cross-country analysis, *Review of Economic Studies* 81, 818–850.

Hosios, Arthur J, 1990, On the Efficiency of Matching and Related Models of Search and Unemployment, *Review of Economic Studies* 57, 279–98.

Hsieh, Chang-Tai, and Peter J. Klenow, 2009, Misallocation and manufacturing tfp in china and india, *The Quarterly Journal of Economics* 124, 1403–1448.

Jovanovic, Boyan, and Balazs Szentes, 2013, On the Market for Venture Capital, *Journal of Political Economy* 121, 493 – 527.

Kaplan, Steven N, and Joshua Rauh, 2013, It's the market: The broad-based rise in the return to top talent, *Journal of Economic Perspectives* 27, 35–56.

Kaplan, Steven N., and Per Stromberg, 2001, Venture Capitals As Principals: Contracting, Screening, and Monitoring, *American Economic Review* 91, 426–430.

Kaplan, Steven N., and Per Stromberg, 2003, Financial contracting theory meets the real world: An empirical analysis of venture capital contracts, *Review of Economic Studies* 70, 281–315.

Karabarbounis, Loukas, and Brent Neiman, 2014, The global decline of the labor share, *The Quarterly journal of economics* 129, 61–103.

Kneer, Christiane, 2013, The Absorption of Talent into Finance: Evidence from U.S. Banking Deregulation, Working Papers 391, De Nederlandsche Bank.

Kuhn, Moritz, Moritz Schularick, and Ulrike I Steins, 2020, Income and wealth inequality in america, 1949–2016, *Journal of Political Economy* 128, 3469–3519.

Kurlat, Pablo, 2019, The social value of financial expertise, *American Economic Review* 109, 556–90.

Levine, Ross, 2005, Finance and growth: Theory and evidence, in Philippe Aghion, and Steven Durlauf, eds., *Handbook of Economic Growth*, chapter 12, 865–934 (Elsevier).

Lucas, Robert Jr., 1988, On the mechanics of economic development, *Journal of Monetary Economics* 22, 3–42.

Lucas, Robert Jr., and Leonard Rapping, 1969, Real Wages, Employment, and Inflation, *Journal of Political Economy* 77, 721–54.

Maddala, G S, and Shaowen Wu, 1999, A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test, *Oxford Bulletin of Economics and Statistics* 61, 631–52.

Mayer-Haug, Katrin, Stuart Read, Jan Brinckmann, Nicholas Dew, and Dietmar Grichnik, 2013, Entrepreneurial talent and venture performance: A meta-analytic investigation of SMEs, *Research Policy* 1–23.

Metrick, Andrew, and Ayako Yasuda, 2010, The economics of private equity funds, *The Review of Financial Studies* 23, 2303–2341.

Midrigan, Virgiliu, and Daniel Yi Xu, 2014, Finance and Misallocation: Evidence from Plant-Level Data, *American Economic Review* 104, 422–58.

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny, 1991, The Allocation of Talent : Implications for Growth, *The Quarterly Journal of Economics* 106, 503–530.

Nash, John, 1950, The bargaining problem, *Econometrica* 18, 155–162.

Nash, John, 1953, Two-person cooperative games, *Econometrica* 21, 128–140.

Pesaran, M. Hashem, 2007, A simple panel unit root test in the presence of cross-section dependence, *Journal of Applied Econometrics* 22, 265–312.

Philippon, Thomas, 2007, Why has the us financial sector grown so much? the role of corporate finance.

Philippon, Thomas, 2010, Financiers versus Engineers: Should the Financial Sector be Taxed or Subsidized?, *American Economic Journal: Macroeconomics* 2, 158–182.

Philippon, Thomas, 2015, Has the us finance industry become less efficient? on the theory and measurement of financial intermediation, *American Economic Review* 105, 1408–38.

Philippon, Thomas, and Ariell Reshef, 2012, Wages and Human Capital in the U.S. Finance Industry: 1909–2006, *The Quarterly Journal of Economics* 127, 1551–1609.

Piketty, Thomas, 2014, *Capital in the twenty-first century* (The Belknap Press of Harvard Univ. Press, Cambridge, Mass. [u.a.]).

Piketty, Thomas, and Gabriel Zucman, 2014, Capital is back: Wealth-income ratios in rich countries, 1700-2010, *The Quarterly Journal of Economics* .

Restuccia, Diego, and Richard Rogerson, 2008, Policy distortions and aggregate productivity with heterogeneous plants, *Review of Economic Dynamics* 11, 707–720.

Romer, Paul M, 1986, Increasing Returns and Long-Run Growth, *The Journal of Political economy* 94, 1002–1037.

Rosen, Sherwin, 1981, The economics of superstars, *The American economic review* 71, 845–858.

Santarelli, Enrico, Luuk Klomp, and A.Roy Thurik, 2006, Gibrat's law: An overview of the empirical literature, in Enrico Santarelli, ed., *Entrepreneurship, Growth, and Innovation*, volume 12 of *International Studies in Entrepreneurship*, 41–73 (Springer US).

Scheidel, Walter, 2017, *The great leveler: violence and the history of inequality from the Stone Age to the twenty-first century* (Princeton University Press).

Shimer, Robert, and Lones Smith, 2001, Matching, search, and heterogeneity, *The B.E. Journal of Macroeconomics* 1, 1–18.

Shu, Pian, 2015, Innovating in Science and Engineering or 'Cashing In' on Wall Street? Evidence on Elite STEM Talent, Working papers, HBS.

Solt, Frederick, 2016, The standardized world income inequality database, *Social Science Quarterly* .

Timmer, Marcel, Gaaitzen de Vries, and Klaas de Vries, 2015, *Patterns of structural Change in Developing Countries*, 65–83, Routledge Handbooks (Routledge).

Wadhwa, Vivek, Ben A. Rissing, and G. Gereffi, 2006, Industry trends in engineering offshoring, report, Duke University.

Watanabe, Makoto, 2010, A model of merchants, *Journal of Economic Theory* 145, 1865–1889.

Wolff, Edward N, 2016, Household wealth trends in the united states, 1962 to 2013: What happened over the great recession?, *RSF: The Russell Sage Foundation Journal of the Social Sciences* 2, 24–43.

Appendix A Static Model

Appendix A.1 Proof of Proposition 1

The proof is organized in the following way. First, I show that as long as γ is greater than 1/2, aggregate output $Y(\gamma)$ decreases with γ. Second, depending on other parameters, $Y(\gamma)$ is either a strictly increasing or strictly decreasing or a non-monotonic function of γ for the whole interval $\gamma \in [0, 1/2]$. In the latter case, I prove that the function is a convex function for $\gamma \in [0, 1/2]$.

For $\gamma > \mathbf{1/2}$, aggregate output $Y(\gamma)$ is given by

$$Y(\gamma) = \frac{(\beta^i - (1-\gamma)\beta^e)}{(1-(1-\gamma)\beta^e)}(1-\beta^e)F_{LH} + \frac{(1-\beta^i)}{1-(1-\gamma)\beta^e}(1-\beta^e)F_{LL} + (1-\gamma)\beta^e F_{HH}. \tag{21}$$

To show whether it is increasing or decreasing, we take the derivative of (21) with respect to γ:

$$\frac{\partial Y}{\partial \gamma}|_{\gamma>1/2} = \frac{\beta^e(1-\beta^i)(1-\beta^e)}{(1-(1-\gamma)\beta^e)^2}F_{LH} - \beta^e F_{HH} - \frac{\beta^e(1-\beta^i)}{(1-(1-\gamma)\beta^e)^2}F_{LL}. \tag{22}$$

We need to estimate the sign of expression (22). Since F is a supermodular function, $F_{LL} \geq 0$ and $F_{HH} \geq F_{LH}$. Hence

$$\frac{\partial Y}{\partial \gamma}|_{\gamma>1/2} \leq \frac{\beta^e(1-\beta^i)(1-\beta^e)}{(1-(1-\gamma)\beta^e)^2} - \beta^e. \tag{23}$$

Expressing the two terms on the right-hand side of condition (23) using a common denominator, then replacing γ in the numerator by 1/2 (this value makes the numerator as small as possible), and finally expanding, we obtain:

$$\frac{\partial Y}{\partial \gamma}|_{\gamma>1/2} \leq \beta^e \frac{(1-\beta^e)(1-\beta^i) - (1-\beta^e+\gamma\beta^e)^2}{(1-(1-\gamma)\beta^e)^2} \leq \beta^e \frac{(1-\beta^e)(1-\beta^i) - (1-\beta^e+0.25(\beta^e)^2)}{(1-(1-\gamma)\beta^e)^2} \leq 0. \tag{24}$$

Here is the end of the proof of the first part. To prove the second part, I follow a similar procedure. I calculate the first derivative and restrict my attention to the case in which the first derivative is neither positive or negative for the whole interval $\gamma \in [0, 1/2]$. Then, I show that in this case the second derivative is positive, i.e. the function is convex.

For $\gamma \in [\mathbf{0}, \mathbf{1/2}]$, the aggregate output $Y(\gamma)$ is given by

$$\begin{aligned} Y(\gamma) = \frac{(\beta^i - \gamma\beta^e)}{(1-\gamma\beta^e)}\left[(1-2\gamma)\beta^e F_{HH} + (1-\beta^e)F_{LH}\right] + \\ \frac{(1-\beta^i)}{(1-\gamma\beta^e)}\left[(1-2\gamma)\beta^e F_{HL} + (1-\beta^e)F_{LL}\right] + \gamma\beta^e F_{HH}. \end{aligned} \tag{25}$$

Calculating the first derivative from (25), we obtain

$$\frac{\partial Y}{\partial \gamma}|_{\gamma\in[0,1/2]} = \frac{1}{(1-\gamma\beta^e)^2}\left[\beta^e F_{HH}\left(1+2\gamma\beta^e-(\gamma\beta^e)^2+\beta^i\beta^e-\beta^e+2\beta^i\right)-\right.$$
$$\left.(1-\beta^e)\beta^e(1-\beta^i)(F_{LH}-F_{LL})-(2-\beta^e)\beta^e(1-\beta^i)F_{HL}\right]. \tag{26}$$

The first derivative is negative for the whole interval $\gamma \in [0, 1/2]$ if

$$\beta^e F_{HH}\left(1+2\gamma\beta^e-(\gamma\beta^e)^2+\beta^i\beta^e-\beta^e+2\beta^i\right)-(2-\beta^e)\beta^e(1-\beta^i)F_{HL} <$$
$$(1-\beta^e)\beta^e(1-\beta^i)(F_{LH}-F_{LL}). \tag{27}$$

The left-hand side of inequality (27) increases with γ, while the right-hand side of inequality (27) is independent of γ. Hence if inequality (27) holds for $\gamma = 1/2$, it holds for any $\gamma \in [0, 1/2]$.

$$\left[\beta^e F_{HH}\left(1-0.25(\beta^e)^2+\beta^i\beta^e+2\beta^i\right)-\right.$$
$$\left.(1-\beta^e)\beta^e(1-\beta^i)(F_{LH}-F_{LL})-(2-\beta^e)\beta^e(1-\beta^i)F_{HL}\right] < 0 \tag{28}$$

If inequality (28) holds, the first derivative is negative. In the opposite case, the sign of the derivative is unknown. Inequality (28) imposes the restriction on the set of exogenous parameters.

We now calculate the second derivative and check its sign:

$$\frac{\partial^2 Y}{\partial \gamma^2}|_{\gamma\in[0,1/2]} = \frac{2\beta^e}{(1-\gamma\beta^e)^3}\left[\beta^e F_{HH}\left(1+\beta^i\beta^e+2\beta^i\right)-\right.$$
$$\left.(1-\beta^e)\beta^e(1-\beta^i)(F_{LH}-F_{LL})-(2-\beta^e)\beta^e(1-\beta^i)F_{HL}\right]. \tag{29}$$

If the second derivative is positive, the function is convex. I need to show that the right-hand side of (29) is positive. As we have seen, when inequality (28) does not hold, the sign of the first derivative is unknown, but it imposes the restriction on the set of exogenous parameters. This is the case in which we need to know the sign of the second derivative. We can estimate the right-hand side using the complementary inequality to condition (28):

$$\beta^e F_{HH}\left(1+\beta^i\beta^e+2\beta^i\right)-(2-\beta^e)\beta^e(1-\beta^i)F_{HL}-(1-\beta^e)\beta^e(1-\beta^i)(F_{LH}-F_{LL}) \geq$$
$$\beta^e F_{HH}0.25(\beta^e)^2 > 0. \tag{30}$$

This completes the proof of the first part. We show that the sign of the first derivative is either negative or unknown. In the case in which it is unknown, we prove that the second derivative is strictly positive. Hence, the solution of the planner's problem can be either 0 or 1/2. □

Appendix A.2 Proof of Proposition 2

The characterization of a decentralized equilibrium is the following triplet: two prices and the share of talented agents in finance (p^i, p^e, γ). We have:

$$\begin{aligned} p^i &\leq (1-\delta)(F_{HH} - d^i(\gamma) - d^e(\gamma)), \\ p^e &\leq \delta(F_{HH} - d^i(\gamma) - d^e(\gamma)), \\ \frac{\min \gamma, 1-\gamma}{\gamma}(p^i + p^e) &= \delta\left(F_{HH} - d^i(\gamma)\right) + (1-\delta)d^e(\gamma). \end{aligned} \tag{31}$$

As a reminder, three types of agents affect the solution of intermediated matching: capital-abundant investors, talented entrepreneurs, and bankers. The number of investors is β^i, the equilibrium number of bankers is $\gamma\beta^e$, and the equilibrium number of entrepreneurs is $(1-\gamma)\beta^e$. There are two markets and consequently two prices that clear them: entrepreneur–banker and investor–banker. System (31) should be solved differently depending on who is on the short side of both markets: capital-abundant investors, talented entrepreneurs, or bankers. I show that a solution exists only if capital-abundant investors are on the long side of the investor–banker market.

Furthermore, the condition $\gamma^* = 1/2$ imposes an additional restriction on the set of exogenous parameters, and eliminates a possible solution with bankers being on the short side with respect to talented entrepreneurs on the entrepreneur–banker market. However, the solution does not always exist. The existence condition is stated, as well.

The system can be solved backward. First, we need to define who is on the short side of the market: capital-abundant investors, talented entrepreneurs, or bankers. Second, I solve the random matching problem for a given size of the financial sector $\gamma\beta^e$ to determine the outside options of capital-abundant investors $d^i(\gamma)$ and talented entrepreneurs $d^e(\gamma)$ in the case in which they decide not to be matched with a high-type counterpart with certainty through a banker. Third, using the solution of random matching as outside options, I solve the intermediated matching problem for capital-abundant investors and talented entrepreneurs.

Capital-abundant investors are on the short side: The number of capital-abundant investors is lower than the number of bankers who provide services for investors $\beta^i < \gamma\beta^e$. Hence, competition among bankers drives the price p^i down to zero. Furthermore, the number of bankers cannot be higher than the number of talented entrepreneurs. Otherwise, bankers' income is zero, and any talented agent strictly prefers to be an entrepreneur. Thus, if capital-abundant investors are on the short side of the investor–banker market, the share of talented agents in finance must be $\gamma \leq 1/2$.

If $\gamma \leq 1/2$ and $p^i = 0$, the system (31) collapses to one condition:

$$d^e(\gamma) = 0. \tag{32}$$

Condition (32) does not hold unless $\delta = 0$. Hence, capital-abundant investors cannot be on the short side in equilibrium.

Capital-abundant investors are on the long side: The number of capital-abundant investors is higher than the number of bankers who provide services for investors $\beta^i \geq \gamma\beta^e$. Hence, bankers push capital-abundant investors to their outside options. The first equation of system (31) becomes an equality. Two cases are possible.

First, if the number of bankers is lower than the number the talented entrepreneurs $\gamma \leq 1/2$, $1 - 2\gamma\beta^e$ talented entrepreneurs are left for random matching. We assume that investors have no access to a storage technology, while entrepreneurs have no opportunity for outside borrowing. Thus, the outside options for a random match—the set D in (7)—are $(0, 0)$. The solution of the Nash bargaining problem gives the value of random matching for capital-abundant investors, which is equal to the probability of matching with somebody $\frac{1-2\gamma\beta^e}{1-\gamma\beta^e}$ multiplied by the sum of two terms: the probability of matching with a talented entrepreneur $\frac{(1-2\gamma)\beta^e}{1-2\gamma\beta^e}$ multiplied by the fraction of the project's output received by the investor $(1-\delta)F_{HH}$; and the probability of matching with an ordinary entrepreneur $\frac{1-\beta^e}{1-2\gamma\beta^e}$ multiplied by the fraction of the project's output received by the investor $(1-\delta)F_{LH}$:

$$d^i = \frac{1-2\gamma\beta^e}{1-\gamma\beta^e}\frac{1-\delta}{1-2\gamma\beta^e}\left[(1-2\gamma)\beta^e F_{HH} + (1-\beta^e)F_{LH}\right]. \tag{33}$$

A similar expression can be obtained for the talented entrepreneur. The probability of matching with somebody for a talented entrepreneur is equal to 1, so

$$d^e = \frac{\delta}{1-\gamma\beta^e}\left[(\beta^i - \gamma\beta^e)F_{HH} + (1-\beta^i)F_{HL}\right].$$

Due to the supermodularity of the output function, sorting is possible. There exists a separating equilibrium such that the incentive compatibility constraint for the capital-scarce investor holds (the low type has no incentive to mimic the high type). The capital-abundant investor is indifferent between being randomly matched and being matched by a banker, while the capital-scarce investor is strictly better off under random matching. In this case, the system (31) takes the form below:

$$\begin{aligned} p^i &= (1-\delta)(F_{HH} - d^i(\gamma) - d^e(\gamma)), \\ p^e &= \delta(F_{HH} - d^i(\gamma) - d^e(\gamma)), \\ (p^i + p^e) &= \delta\left(F_{HH} - d^i(\gamma)\right) + (1-\delta)d^e(\gamma). \end{aligned} \tag{34}$$

Surprisingly, as long as $\gamma \leq 1/2$, the income of a banker is an increasing function of the number of bankers, while the income of an entrepreneur is a decreasing function of the number of bankers. The rise of bargaining power δ has no effect on the banker's income and a positive one on entrepreneurial income. The solution of the system (34) is linear in δ:

$$\tilde{\gamma} = \delta\left[\frac{1}{\beta^e} + \frac{1-\beta^e}{\beta^e}\frac{F_{HH}-F_{LH}}{F_{HH}} - \frac{2(1-\beta^i)}{\beta^e}\frac{F_{HH}-F_{HL}}{F_{HH}}\right] - \frac{1-\beta^e}{\beta^e}\frac{F_{HH}-F_{LH}}{F_{HH}}. \tag{35}$$

There exist two thresholds $\bar{\delta} > 0$, such that $\tilde{\gamma} = 0$, and $\tilde{\delta} > 0$, such that $\tilde{\gamma} = 1/2$:

$$\bar{\delta} = \frac{(1-\beta^e)(F_{HH}-F_{LH})}{(1-\beta^e)(F_{HH}-F_{LH}) + (2\beta^i - 1)(F_{HH}-F_{HL}) + F_{HL}}, \tag{36}$$

$$\tilde{\delta} = \frac{(1-\beta^e/2)F_{HH} - (1-\beta^e)F_{LH}}{(1-\beta^e)(F_{HH}-F_{LH}) + (2\beta^i - 1)(F_{HH}-F_{HL}) + F_{HL}}. \tag{37}$$

Depending on parameter values, both $\bar{\delta}$ and $\tilde{\delta}$ can potentially be greater than 1. The solution $\tilde{\gamma}$ exists only for $\delta \in [\min[\bar{\delta}, 1], \min[\tilde{\delta}, 1]]$. The solution $\tilde{\gamma}$ exists as long as $\bar{\delta} \leq 1$. Using expression (36), the latter can be rewritten as follows:

$$(2\beta^i - 1)(F_{HH} - F_{HL}) + F_{HL} \geq 0. \tag{38}$$

Second, the number of bankers is greater than or equal to the number of talented entrepreneurs $\gamma \geq 1/2$. Thus, all talented entrepreneurs are matched by bankers. The number of capital-abundant investors $\beta^e(1-\gamma)$ are matched by bankers. In this case, the solution of the Nash bargaining problem for random matching is given by

$$d^i = \frac{(1-\beta^e)}{1-\beta^e(1-\gamma)}(1-\delta)F_{LH}, \tag{39}$$

$$d^e = \frac{\delta}{1-\beta^e+\gamma\beta^e}\left[(\beta^i - \beta^e + \gamma\beta^e)F_{HH} + (1-\beta^i)F_{HL}\right]. \tag{40}$$

If the number of bankers is greater than the number of talented entrepreneurs, competition among bankers drives the price p^i down to zero. In this case, system

(31) takes the form below:

$$\begin{array}{c} p^i = (1-\delta)(F_{HH} - d^i(\gamma) - d^e(\gamma)), \\ p^e = 0, \\ \frac{1-\gamma}{\gamma} p^i = \delta\left(F_{HH} - d^i(\gamma)\right) + (1-\delta) d^e(\gamma). \end{array} \tag{41}$$

As we can see, the banker's income is a decreasing function of the bargaining power δ, while entrepreneurial income is an increasing function of δ. Furthermore, the expected income of a banker grows with γ. System (41) can be expressed in the form of a quadratic equation in γ:

$$\begin{array}{c} \gamma^2 + \left[(1-\delta)\frac{(1-\beta^e)}{\beta^e}\frac{F_{HH}-F_{LH}}{F_{HH}} + \frac{\delta-\beta^e}{\beta^e} + \delta(1-\delta)\right]\gamma - \\ -(1-\delta)^2\frac{(1-\beta^e)}{\beta^e}\frac{F_{HH}-F_{LH}}{F_{HH}} - (1-\beta^i)(1-\delta)\frac{\delta}{\beta^e}\frac{F_{HH}-F_{HL}}{F_{HH}} = 0. \end{array} \tag{42}$$

The solution of quadratic equation (42) contains two roots, but one root is always negative. For the solution to exist, the second root has to be greater than 1/2. Let $\hat{\gamma}$ be the positive solution of (42). This solution exists as long as

$$\frac{\delta(1+\beta^e(1-\delta))}{4(1-\delta)} \leq (1-\beta^i)\delta\frac{F_{HH}-F_{HL}}{F_{HH}} + (1-\beta^e)\frac{F_{HH}-F_{LH}}{F_{HH}}(1/2-\delta). \tag{43}$$

Analyzing condition (43), we can conclude that the condition is likely to be satisfied when: the dispersion of wealth k^H/k^L is high; the share of capital-abundant investors β^i is low; and the bargaining power of entrepreneurs δ is relatively low. Furthermore, if $\delta \leq 1/2$, condition (43) is likely to be satisfied when the dispersion of talent is high and the share of talented agents β^e is low. When condition (43) is satisfied with equality, it can be rewritten as the definition of $\hat{\delta}$ defined in Proposition 2.

The constrained efficient allocation is $\gamma^* = 1/2$: This implies that $\mathbf{\Delta Y}$ given by expression (5) is positive and can be rewritten in the following form:

$$(2\beta^i - 1)(F_{HH} - F_{HL}) + F_{HL} < -\frac{2(1-\beta^i)(1-\beta^e)}{2-\beta^e}(F_{LH} - F_{LL}). \tag{44}$$

The right-hand side of inequality (44) is negative, therefore its left-hand side is negative as well. If we compare the left-hand side of (44) with the left-hand side side of expression (38), they are exactly the same. Hence, if $\gamma^* = 1/2$, the solution $\tilde{\gamma}$ does not exist. □

Appendix A.3 The solution of the decentralized equilibrium if $\gamma^* = 0$

In this section, I show the solution of the decentralized equilibrium in the case in which the constrained efficient allocation is 0. The proposition below summarizes the case:

PROPOSITION 3. *If the constrained efficient allocation is $\gamma^* = 0$, then both equilibria with few $\tilde{\gamma}$ and many $\hat{\gamma}$ bankers are possible. In this case, there is a range of $\delta \in [\tilde{\delta}, 1]$, such that the decentralized equilibrium is constrained efficient.*

Proof: As shown in appendix Appendix A.2, the solution $\hat{\gamma}$ exists as long as condition (43) holds, and it does not depend on whether the constrained efficient allocation is 0 or 1/2; the solution $\tilde{\gamma}$ does not exist if $\gamma^* = 1/2$. The solution $\tilde{\gamma}$ exists as long as $\bar{\delta} < 1$, defined by expression (36). The latter is likely to be satisfied if $\gamma^* = 0$. □

Appendix A.4 A toy example

The simple example from table 2 shows the disparity between the first best and random matching: the loss of aggregate output due to the misallocation of capital caused by private information in this economy can be severe. I consider the case in which the production function is simply the product of two inputs $F = zk$, and the value of the high type is one with a probability of one-quarter, while the value of the low type is zero with the complementary probability for both the distribution of talent and the distribution of wealth. Hence, only if two high types are matched is any output (one unit) produced. It happens with probability 1/16 in the case of random matching and with probability 1/4 in the case of assortative matching (the first best). Table 2 summarizes the information described above. As we can

Table 2. A simple example

	value	probability
z^H	1	1/4
z^L	0	3/4
k^H	1	1/4
k^L	0	3/4
Random matching	1/16	
Assortative matching	1/4	

see, output is four times lower in the case of random matching compared to the first best due to capital misallocation. This brings us to the first main question of whether the financial sector can mitigate this capital misallocation.

Based on the simple example in Table 2 we can calculate aggregate output in the constrained efficient case, we obtain $1/2\beta^e F_{HH} = 1/8$, which is twice as large as in the case of random matching (the economy without finance), but still two times lower than in the first best. In the case of the simple example, we can say that the financial sector undoes half of the financial friction.

Appendix A.5 Taxation in the static model

The question is whether it is possible to restore efficiency. The answer is yes. As discussed, there is a restriction on parameters that restores efficiency. If a policy instrument can be introduced that may directly affect one of the exogenous parameters, it would be easy to ensure efficiency in the model. For example, if the planner could set the bargaining power of entrepreneurs to the particular value of $\hat{\delta}$, it would make the decentralized equilibrium efficient. However, it is not very intuitive to think that such policies exist.

The more interesting question is whether it is possible to restore efficiency using only one tax instrument. Fixing the set of parameters to values such that the decentralized equilibrium exists and is inefficient, I take the tax on the financial sector to be the available tax instrument.

The issue in this economy is that the return to finance is too high in comparison with entrepreneurship. Hence an efficient policy should decrease the return to finance and/or increase the return to entrepreneurship. The former can be done through taxation of the financial sector. The latter can be done through subsidizing entrepreneurship. Taxation of the financial sector has been a hot topic since the Great Recession, especially in the European Union.[8] Subsidies for entrepreneurship are quite common: governments and donors spend billions of dollars subsidizing entrepreneurship training programs around the world (see, for example, Santarelli et al. (2006)).

I show how a tax τ on bankers' incomes can work. The revenue from this tax is distributed by lump-sum transfers T to balance the government's budget. The last equation of system (45) represents the government's budget constraint. The

[8]See the discussion of taxation proposals at the European Commission web page: `http://ec.europa.eu/taxation_customs/taxation/other_taxes/financial_sector/index_en.htm`.

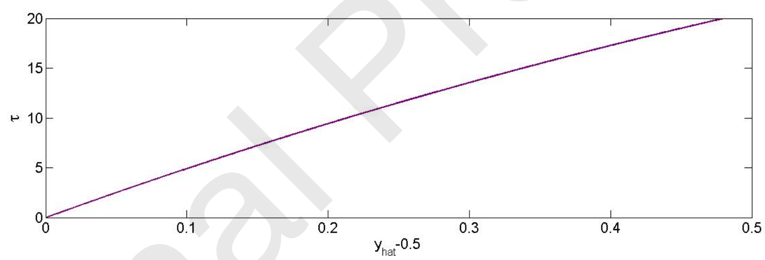
system below characterizes the equilibrium with taxation:

$$\begin{aligned} x^e &= (1-\delta)(F_{HH} - d^i(\gamma) - d^e(\gamma)) + T, \\ c &= (1-\delta)(F_{HH} - d^i(\gamma) - d^e(\gamma)) - 2(1-\delta)T - \tau, \\ x^e &= \tfrac{1-y}{y}c, \\ T &= \gamma\beta^e\tau. \end{aligned} \tag{45}$$

Given the constrained efficient level $\gamma^* = 1/2$, I impose that $\gamma = \gamma^*$ and calculate the corresponding tax rate. The solution of the system can be represented graphically. Figure 6 plots the tax on banking income in percent as a function of the distortion (inefficiency) $\hat{\gamma} - \gamma^*$. The optimal tax is zero when there is no distortion, and increases with the size of the distortion as expected. The closed-form solution of the system defining the tax on banking income as a function of all exogenous parameters is:

$$\tau = \frac{2\delta(1-\delta)\beta^e F_{HH}}{(2-\beta^e)}\left[\frac{2(1-\beta^i)}{\beta^e}\frac{F_{HH}-F_{HL}}{F_{HH}} + \frac{1-\beta^e}{\beta^e}(1 - 2\delta\frac{F_{HH}-F_{LH}}{F_{HH}} - 1 - \frac{2\delta-\beta^e}{2\delta\beta^e(1-\delta)}\right].$$

Figure 6. Tax on financial income vs. inefficiency

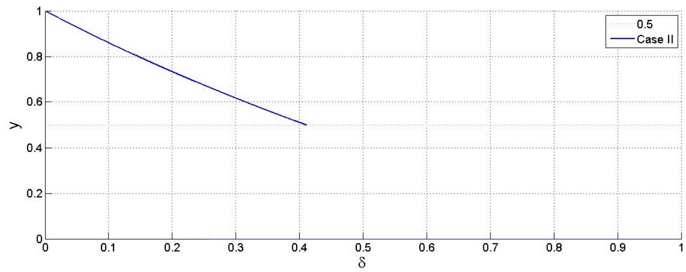


Appendix A.6 Comparative statics

Returning to the solution of the decentralized equilibrium, I analyze the comparative statics of the outcome of the model as exogenous parameters change. The decentralized equilibrium is a function of all exogenous parameters: $\hat{\gamma} = f(\delta, \beta^e, z^H/z^L, \beta^i, k^H/k^L)$. For example, Figure 7 presents the solution $\hat{\gamma}$ as a function of the bargaining power δ. As we can see, the decentralized equilibrium exists only for $\delta \in [0, \hat{\delta}]$; there is no solution for $\delta > \hat{\delta}$. The decentralized equilibrium coincides with the constrained efficient outcome only for one particular value of the bargaining power $\hat{\delta}$.

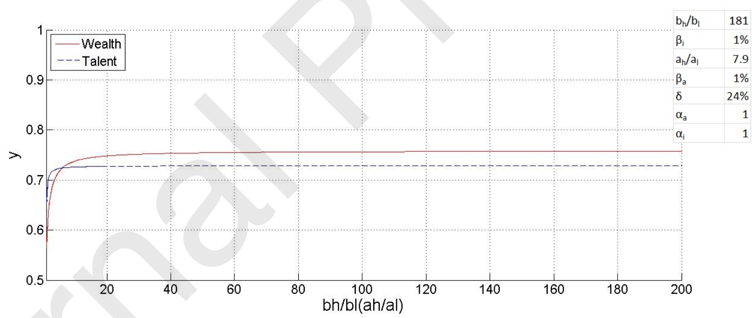
Figure 8 presents the solution of the decentralized equilibrium as a function of wealth k^H/k^L and talent z^H/z^L dispersion. As we can see, wealth dispersion

Figure 7. Fraction of bankers vs. bargaining power of entrepreneur (efficient fraction is 1/2)



has a stronger impact on the size of the financial sector. More importantly, the static model predicts that an increase in wealth inequality will be associated with the growth of finance. When the rich get richer, they demand more finance. This is in line with empirical evidence. However, the wealth distribution has been considered completely exogenous up until now. The next section endogenizes the wealth distribution by introducing dynamics into the model.

Figure 8. Fraction of bankers vs. dispersion of wealth (talent) (efficient fraction is 1/2)



Appendix A.7 Talent in entrepreneurship and finance

I assume that talented entrepreneurs are better in both producing and screening, so only talented entrepreneurs would be selected to become bankers to improve the matching efficiency in the economy, at the opportunity cost of giving up their talent in entrepreneurial activities. In other words, the talent at entrepreneurship and finance is perfectly correlated. In this subsection, I argue that the case of perfect correlation is empirically relevant, and it is what makes the tradeoff in the model interesting. However, one can easily relax the assumption of perfectly correlated abilities in finance and entrepreneurship. By assuming that the abilities in two industries are drawn sequentially, It is possible to accommodate arbitrary correlation and yet keep the binomial distribution of abilities in both industries (either high or low). The talent is a scarce resource in both finance and entrepreneurship. For exposition purposes, the total number of talented agents is restricted to be the same in both industries. The entrepreneurial abilities are drawn first, let β^e be the fraction of talented entrepreneurs, and $1-\beta^e$ the fraction of ordinary entrepreneurs. Next, let ψ_H be the probability of being also a talented banker conditional on being a talented entrepreneur, while ψ_L be the probability of being a talented banker conditional on being a ordinary entrepreneur. In order to reduce the numbers of parameters to one, I impose that $\psi_H = 1 - \psi_L = \psi$. Hence, the fraction $\beta^b_H = \beta^e\psi$ ($\beta^b_L = (1-\beta^e)(1-\psi)$) is the fraction of talented bankers, who are also talented (ordinary) entrepreneurs. Then, by varying ψ, any correlation can be spanned. For example, the baseline case of perfect correlation is $\psi = 1$, or the perfect negative correlation case is $\psi = 0$. For $\psi \in (0,1)$, we have two additional cases to consider, which are absent in the case of perfect correlation: a talented entrepreneur, but an ordinary banker; a talented banker but an ordinary entrepreneur. The equilibrium in the perfectly correlated case was constructed in the way that ordinary agents in both industries strictly prefer entrepreneurship, while talented agents in both industries are indifferent between two occupations.

Given that, a talented entrepreneur, but an ordinary banker would strictly prefer entrepreneurship, while a talented banker but an ordinary entrepreneur would strictly prefer finance. Hence, the relevant occupational choice would remain only for the subset of agents that are talented in both industries. This extension introduces a non-zero number of talented agents, who strictly prefer to be one of the two occupations, and consequently might help to match the level of financial employment.

I cite a few selected papers in the introduction, but there is a large empirical literature that estimates the impact of the potential brain drain from industry to finance. Many papers provide indirect empirical evidence on the misallocation of talent. Data from college graduates in the US suggests that the financial sector has become one of the most popular destinations for graduates of elite universities with high levels of raw academic talent, regardless of their major (see Goldin and Katz (2008) for Harvard graduates and Shu (2015) for MIT graduates), and Wadhwa et al. (2006) for Engineering Management graduates at Duke University). Goldin and Katz (2008) calculate that the percentage of male Harvard graduates with occupations in the financial sector 15 years after their graduation almost doubled form the 1970 to 1990 cohort. In addition, Wadhwa et al. (2006) reports that 30 to 40 percent of Duke Masters of Engineering Management were accepting jobs outside of the engineering profession, choosing to become investment bankers or management consultants rather than engineers. More recently, Shu (2015), studying the career choices of MIT graduates, concludes that careers in finance attract students with high raw academic talent. However, it is unclear whether limiting entry into finance due to the financial crisis has improved the overall efficiency in talent allocation. Even though the job opportunities in finance have shrunk dramatically aftermath of the global financial crisis, which we observe in the data, the share of Yale graduates in finance has shown a slightly different pattern: the dot-com crisis seems to have a higher negative impact on Yale graduates career perspectives in finance. Of the 2010 Yale graduates who were working a year out, 14 percent were in business/finance jobs, down from a peak of 22 percent in 2000. Beverly Waters, Office of Institutional Research, Yale University (2011). According to the Princeton Office of Career Services, 35.9 percent went into finance in 2010, down from a peak of 46 percent in 2006. According to the Harvard Office of Career Services, Harvard graduates entering jobs were more likely to enter finance than any other career: 17 percent of new graduates did so. Nevertheless, this share is still significantly lower than it was at the peak in 2008 when 28 percent of employed new graduates worked in finance. Despite this drop, the long term trend is still upwards. The Princeton data shows a sign of recovery to pre-crisis levels.

Kneer (2013) finds that US banking deregulation reduces labor productivity disproportionately in relatively skill-intensive industries. Finally, the McKinsey Global Institute estimates in 2011 that the United States may face a shortfall of almost two million technical and analytic workers over the next ten years. In the aerospace sector, 60 percent of the aerospace workforce is over 45 years old compared with 40 percent in the overall economy. Célérier and Vallée (2019) document that an increasing fraction of graduates from elite universities has been taking up jobs in finance over the recent decades. Besides, the evidence of the brain drain is not limited to the US. Focusing on French engineering graduates, Célérier and Vallée (2019) show that the increase in relative pay has been particularly pronounced for finance workers who have graduated from the very top engineering programs. Based on Swedish administrative data, Böhm et al. (2018) find no evidence that talent in finance improved drain, but show that finance workers capture rising rents over time.

Several studies show that rising compensation levels in finance can explain a significant part of the growth in top incomes for many countries (Guvenen et al., 2014; Kaplan and Rauh, 2013). This empirical pattern motivates the development of theoretical literature on the competition for talent in the financial sector (Cooley et al., 2020; Acharya et al., 2016).

Appendix A.8 Discussion of assumptions

In this section, we review the impact of the different assumptions on the outcome of the model: the inefficiency result and the inequality result. First, the inefficiency result states that the decentralized equilibrium is generically inefficient. Second, the inequality result states that the endogenous growth of wealth inequality leads to the expansion of the financial sector.

Preferences and technology: First, for simplicity, I assume the Cobb–Douglas production function $z^{\alpha_z} k^{\alpha_k}$, which satisfies the supermodularity condition. However, the choice of a production function should not affect the results, because as long as z and k are not fully substitutable, the supermodularity condition holds. According to the Becker–Brock theorem, the supermodularity condition implies that positive assortative matching is the first best allocation of the model. Hence, the results of the model remain unchanged so long as the production function is supermodular. Second, labor could easily be included as an additional input, but this would not add further insights into the questions addressed in this paper. It should not affect the choice of talented agents, but it might have interesting implications for ordinary agents. Third, if we consider a risk-averse utility function instead of a risk-neutral one, all agents would prefer to engage in risk-sharing. If profit sharing and fund pooling are available options, the introduction of risk aversion does not change anything, because expected and realized incomes are the same. If these options are not available, the impact of risk aversion is ambiguous. On the one hand, investors are willing to pay a higher price for intermediated matching. The higher the price there is, the higher the income of a banker there is. On the other hand, due to higher uncertainty about the banker income, risk aversion makes a banking career a less attractive option.

Distribution of types: First, as long as within each period, the wealth distribution is independent of the talent distribution, the investment decision is independent of the occupational choice. This makes the solution of the problem tractable. The consideration of the two-dimensional joint distribution of wealth and talent complicates the analysis enormously without much additional insight into this particular question. Second, the fact that the constrained efficient allocation admits only two values is an artifact of the discrete distribution of talent and the particular type of information advantage for talented agents in finance: a banker with ability z can distinguish between ideas with productivity z and $z' < z$. As long as both assumptions hold, the constrained efficient allocation admits two values (zero and one-half) for each type of talent: the planner would find it optimal either to keep the allocation under random matching or to make it as close as possible to the allocation under assortative matching by exhausting the opportunities for intermediated matching fully. The allocation in the case of a continuous talent distribution would strongly depend on the assumption made with respect to the impact of talent on agents' productivity in the two sectors. Third, intuitively, for the case of continuous wealth distribution, the constrained efficient solution either has a positive share for all values of talent distribution z or there exists a threshold in terms of ability $\bar{z}$, that separates bankers and entrepreneurs. Calculating the decentralized equilibrium is a complicated numerical task.

Different types of frictions: First, this paper focuses on how the financial sector arises as a result of one type of relevant friction, adverse selection. The financial sector clearly provides other useful functions to the economy: it allocates not only information but also decision power and risk. On the theoretical side, the literature considers the following functions of the financial sector: screening to mitigate the effect of adverse selection, monitoring to prevent the effects of moral hazard; auditing, and punishment to mitigate the effects of opportunistic behavior in the context of costly state verification. While Bolton et al. (2016) study moral hazard and the financial sector as a liquidity provider, I consider adverse selection and the financial sector as a classical intermediary. However, we both obtain a similar result in terms of efficiency, but the mechanisms are substantially different. This similarity suggests that the misallocation result might be a general feature of models with financial frictions. Under the assumption that talent in finance affects the efficiency of monitoring, the inequality result is likely to survive as well.

Second, the matching friction is essential for the inefficiency result because, in a perfectly competitive market, prices would take into account the negative externality which arose from the occupational choice. However, more a general form of matching friction, many-to-one matching can be easily introduced into the environment in at least two ways: through diminishing returns on capital and fixed costs of engaging with investors; or through making entrepreneurs' bargaining power depend positively on the number of investors in the market. Ceteris paribus, it is likely that many-to-one matching would lead to more inefficiency in comparison to one-to-one matching. More investors that can be matched with one entrepreneur mean fewer bankers are needed to restore efficiency. The income of a banker increases with the number of investors matched with one entrepreneur. Hence, an even larger fraction of talented agents is attracted to finance. The inefficiency should increase due to both a decline in the constrained efficient fraction of talented agents in finance and a rise in the decentralized one. This mechanism is present in the dynamic part of the paper (section 2).

Third, the issue of competition has been studied extensively. A monopoly is usually viewed as a bad thing. However, in the framework, one monopolistic firm in the financial sector might restore efficiency because it maximizes the total surplus by pushing all agents to their outside option. The monopolist is always on the short side of the market. It would set the prices for its services to make both entrepreneurs and investors indifferent between paying for the services and being matched and being randomly matched for free. On top of this, the monopolist can set wages for its workers (bankers) to make them indifferent between the two sectors. Hence, the monopolist could extract the total surplus and would hire an efficient number of bankers. However, this possible advantage of a monopoly due to information provision does not overcome the general disadvantages of monopoly for society.

The informational friction can be undone without the financial sector, which is the case when the type of investors and the outcome of the match are publically observable. Hence, the entrepreneurs could signal the type by writing a contract conditional on the output and the investor type. In order to make the problem interesting, it is sufficient to make sure that the entrepreneurs could signal the type, which is the case when the types of entrepreneurs are not publicly observable and if any of the following assumptions hold. First, the outcome of the match is not observable or contactable. The justification can be that the outcome depends not only on talent and capital but also on some shock. Second, entrepreneurs do not know the wealth of investors they are dealing with. Even though the latter assumption seems questionable at first, in the venture capital industry it is common for entrepreneurs to be imperfectly informed about the total wealth of investors.[9]

Appendix B Dynamic Model

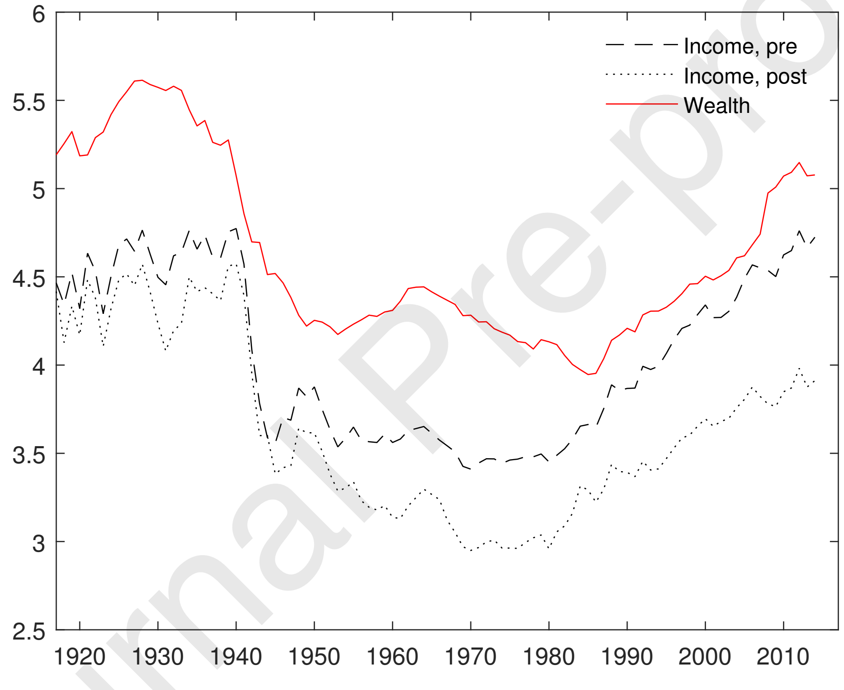
Appendix B.1 US Data

Figure 9 presents the dynamics of income and wealth inequality in the United States. The inequality is measured as the ratio of the top 10 % income(wealth) to the average income (wealth), which corresponds to the ratio of $\frac{k^H}{k^L}$ of the model. All three measures exhibit similar dynamics: the sharp drop during WWII, the

[9]In the case of venture capital, after engaging with a venture capitalist, the entrepreneur faces a substantial degree of uncertainty about the total amount of investment, because of staging. Staging is one of the central incentive mechanisms used in the venture capital industry As shown by Bienz and Hirsch (2011), staging is frequently implemented through multiple negotiated financing rounds. Furthermore, the venture capital literature often assumes that *neither* the inputs of the investor *nor* those of the entrepreneur are contractible. The standard feasible contract in the venture capital literature specifies only a sharing rule and an initial investment, but not the total investment, which, like entrepreneurial inputs, is assumed to be non-contractible. Second, even if the wealth of an investor is observable, the exact amount the investor is willing to stake in a particular project, is likely to be unknown to the entrepreneur.

stagnate inequality from the 1950s to 1970s, the rapid growth from the 1980s onwards. I target income inequality instead of wealth inequality for two reasons. First, the dynamics of the two measures are similar. Through the lens of the model, there are no differences between the two: the income inequality among the young becomes the wealth inequality among the old. Second, focusing on income inequality instead of wealth inequality allows us to study the impact of taxation because Piketty and Zucman (2014) provides the data for pre- and post-tax income inequality.

Figure 9. US Inequality

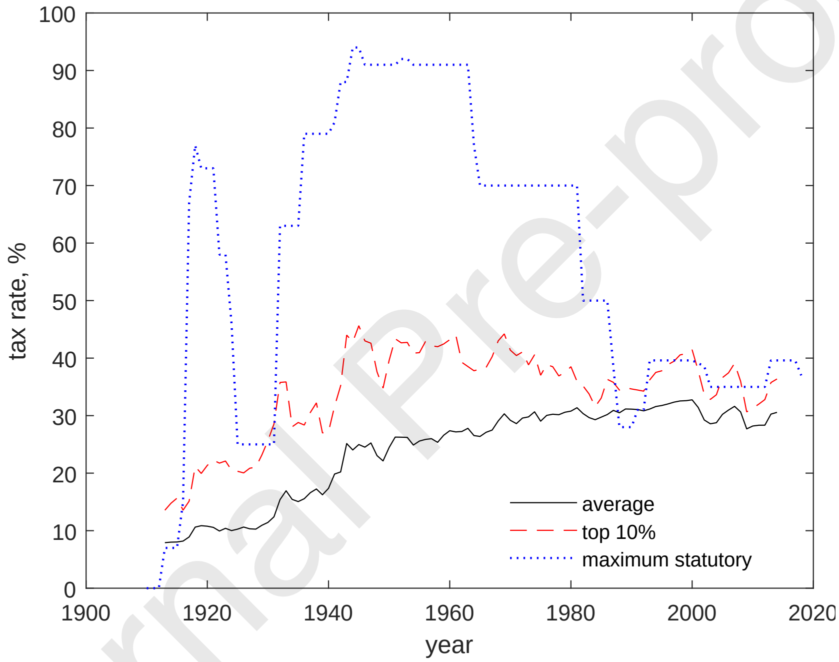


Notes: This figure plots the pre- and post-income inequality as well as wealth inequality. Inequity is measured as the ration between the average income of top 10% and the overall average, which corresponds to k^H/k^L. Data are from Alvaredo et al. (2015) for the period 1917-2015.

Figure 10 presents the evolution of tax rates in the US. The average effective tax rate is the solid black line. The red dashed line is the average tax rate for the top 10%. The maximum statutory tax rate is the dotted blue line. The tax progressivity of the US tax system has declined due to the increase in the average tax rate and the reduction in the top 10% tax rate.

Following Philippon (2007), I define three subindustries within the financial sector: "Banking", "Insurance" and "New Finance". Banks, thrift, and saving

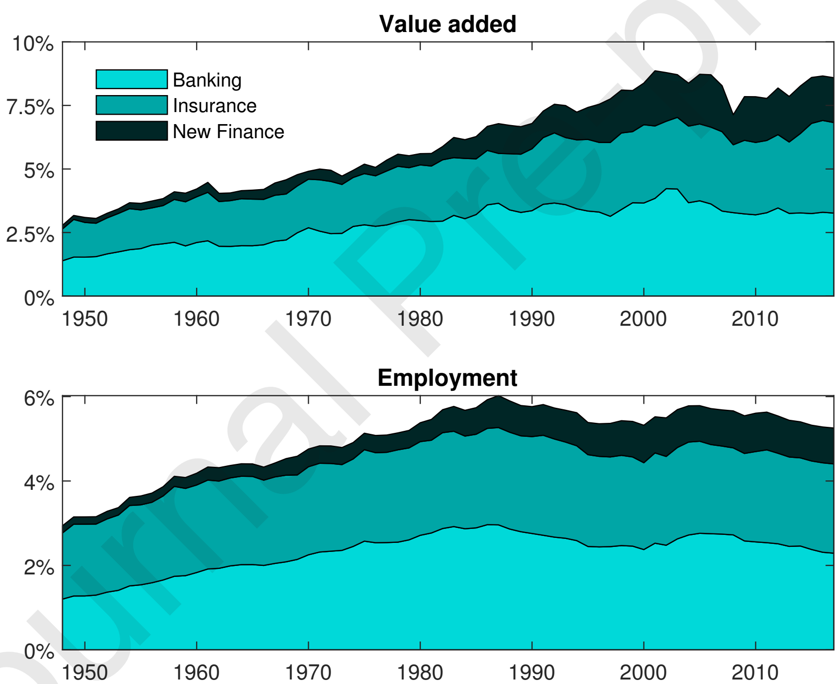
Figure 10. US tax rates



Notes: This figure plots the average tax rate, the average tax rate for the top 10%, and the highest marginal income statutory tax rates. Data are from Alvaredo et al. (2015) and IRS for the period 1909-2018.

institutions are included in "Banking". Securities, commodities, investment offices, funds, trusts, and other financial vehicles, as well as investment banks and private equity, are all included in "New Finance". Figure 11 presents the result of the decomposition. The top panel of Figure 11 shows the size of each subindustry in the percentage of GDP in terms of value-added. All three subindustries have been growing. While the share of banking and insurance increased by less than trice, the share of new finance grew 13 times, reaching almost 2 % of GDP. The bottom panel of Figure 11 shows the size of each subindustry in the percentage of employment. While the employment share of banking and insurance reached a peak in the late 1980s, the employment share of new finance has increased five times, reached almost 1 % of total employment, and continues to grow.

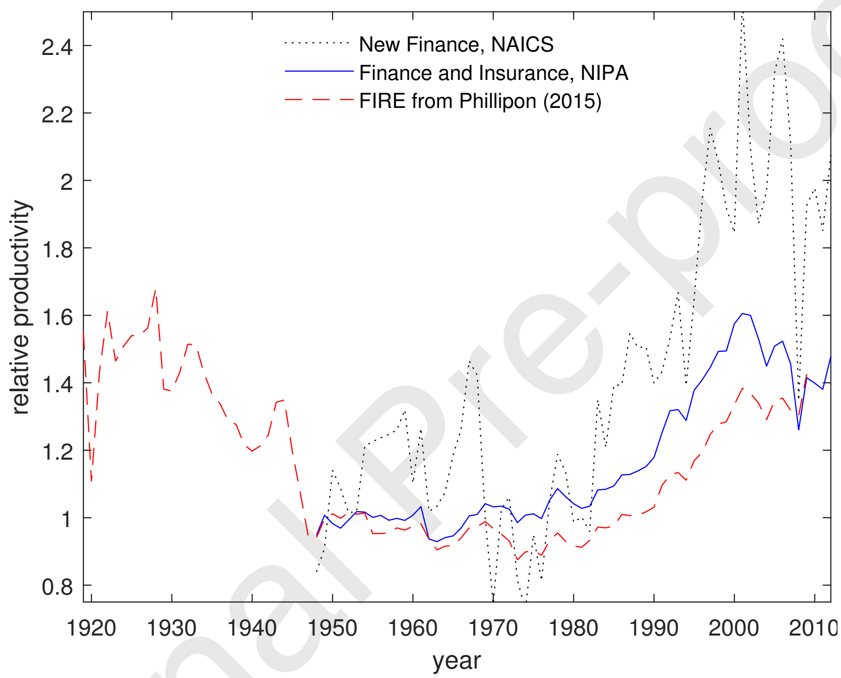
Figure 11. Financial sector in the US



Notes: This figure plots the dynamics of subindustries within the financial sector: "Banking", "Insurance" and "New Finance". Data are form BEA for Value Added and Full-Time and Part-Time Employees by Industry for the period 1947-2017.

This rapid expansion of the GDP shares of finance translates into the substantial increase of the relative productivity of finance, which is defined as the ratio of the GDP share to the employment share of finance. The black doted line is the relative labor productivity of new finance. The red dashed line is the ratio of time series from Figure 1 of the main text. As you can see, the relative productivity of new finance increased by 100% compared to 50 % for finance. Most of the increase occurred starting from the 1980s. This is consistent with one of the main findings of Philippon and Reshef (2012), who document a steep increase of 70 % in relative wages and skill intensity (measured as the relative fraction of college-educated workers), and job complexity.

Figure 12. Finance Relative Productivity



Notes: Figure plots the relative labor productivity in finance defined as the ratio of the GDP share to the employment share of finance. The red dashed line is based on the GDP share from Philippon (2015) and the employment share from Buera and Kaboski (2012) for finance insurance and real estate (FIRE). The blue line is based on more recent data form BEA for Value Added and Full-Time and Part-Time Employees by Industry.

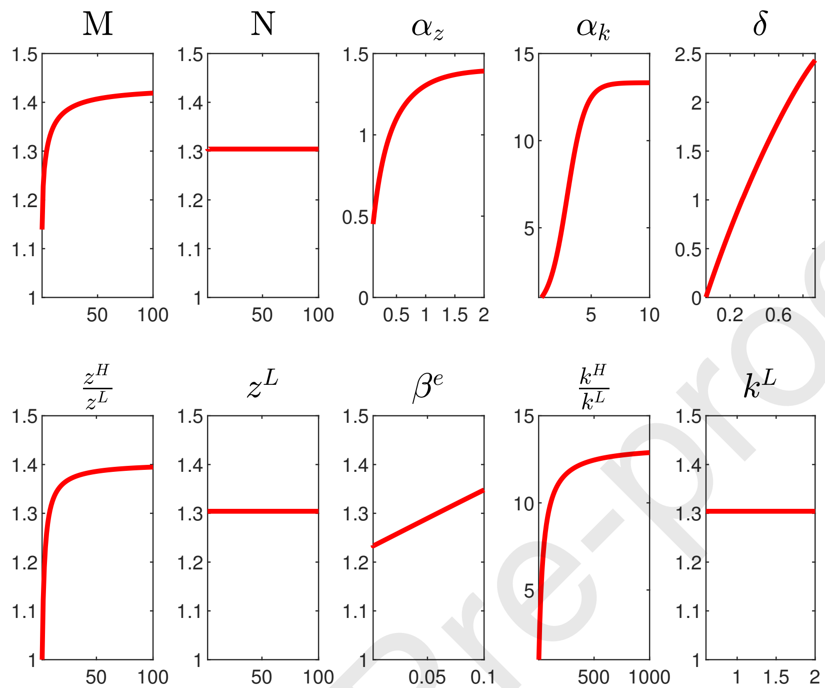
Appendix B.2 Comparative statics

We now consider the solution of the static model for different parameter values. We first fix the parameters to the average values of the baseline estimation as in Model 1 of Table 1, which corresponds to the value of the relative finance productivity 1.32. Figure 13 presents the comparative statics for the relative financial productivity by varying one by one each of all eight parameters.

We first discuss the economic interpretation of parameters. There are eight exogenous parameters: β^e the share of talented agents in the economy; the relative productivity of talented agents in entrepreneurship $\left(\frac{z^H}{z^L}\right)$; the productivity of ordinary entrepreneurs z^L; the capacity of bankers M and N (number of investors and entrepreneurs which can be matched by a banker); δ the bargaining power of entrepreneurs, which determines the surplus split; α^k return to scale on capital; α^z return to scale on talent. Some of these parameters can be time-varying. First, β^e might capture an increase in the supply of high skills. Since the size of the financial sector in terms of employment is limited by the number of talented agents β^e, by increasing β^e, we can match the increase in financial employment. However, β^e also controls the top percentile of the income distribution for investors. I keep β^e constant in all calibrations in order to match the top percentile. Second, the increase of relative productivity of talented agents in entrepreneurship z^H/z^L can be viewed as skill-biased technological change (Acemoglu, 2002). Third, the capacity of bankers M and N is a measure of the banker's productivity, which plays an important role. The increase in capacity captures the superstar effect in the spirit of Rosen (1981). Fourth, δ determines the surplus split and, consequently, the capital share of the output. By reducing δ, we could capture the increase of the capital share, a well-documented phenomenon (Karabarbounis and Neiman, 2014). Fifth, Piketty (2014) argues that diminishing returns on capital, although undoubtedly present, are unlikely to be very strong. A way to capture this idea is to set $\alpha_k \gtrsim 1$.

By varying parameters, we can obtain the variation in the relative financial productivity summarize our results as follows (Figure 13). First, when the banker capacity in terms of investors M increases, the banker's income increases, but the number of bankers declines in order to satisfy the occupational choice condition. Second, the banker capacity in terms of the number of entrepreneurs N does not affect the relative productivity of finance, since, in the proposed equilibrium with many bankers $\gamma > \frac{1}{N}$, the number of entrepreneurs affects neither the income of bankers nor the number of bankers. You can clearly see it from Equation 17. Third, a rise of the relative dispersion of talent $\frac{z^H}{z^L}$ or capital $\frac{k^H}{k^L}$ increases the surplus of intermediated matching. This leads to the increase in the banker income and the number of bankers, which is similar to the simple static model (Figure 8). However, the introduction of banker capacities, in particular in terms of the number of investors, makes the banker income change significantly with the dispersion of capital. the returns to scale on productivity and capital α^z and α^k play a similar role to the relative dispersion of talent $\frac{z^H}{z^L}$ and capital $\frac{k^H}{k^L}$. Forth, the productivity of ordinary entrepreneurs z^L and the capital level of ordinary investors k^L has no impact on the relative financial productivity since it

Figure 13. Comparative Statics



Note: The figure presents the comparative statics for the relative financial productivity by varying one by one each of all eight parameters around the baseline calibration.

plays a role of the numeraire. What is more important to determine the banker income and the number of bankers is the relative surplus of intermediated matching, which is affected neither by the capital level of ordinary investors nor by the productivity of ordinary entrepreneurs. Fifth, the relative productivity increases with the bargaining power of entrepreneurs δ. When $\delta = 0$ the income of entrepreneurs is zero and consequently the income of bankers. Furthermore, all talented agents are bankers. Hence, the relative productivity of finance is zero, because the share of finance in output is zero, while there is a positive fraction of agents working in finance. As δ increases, the income of talented rises, but the number of bankers falls. This leads to an increase in relative financial productivity. Sixth, the relative productivity increases with the fraction of talented agents β^e because it reduces the rationing of bankers in the equilibrium.

Appendix B.3 Calibration Extensions

The goal of this subsection is to see how we can improve upon the baseline calibration if we add more degrees of freedom. In the baseline calibration, M_t is calibrated to match the relative productivity of finance. Another way to change the banker income is through the entrepreneurial bargaining power δ, which affects the share of the surplus, which goes to investors and eventually to bankers. The relative productivity of talented entrepreneurs $\frac{z^H}{z^L}$ affects the surplus of intermediated matching and ensures that the level of inequality is consistent with the data. The return to scale on capital α_t^k controls the speed of wealth growth.

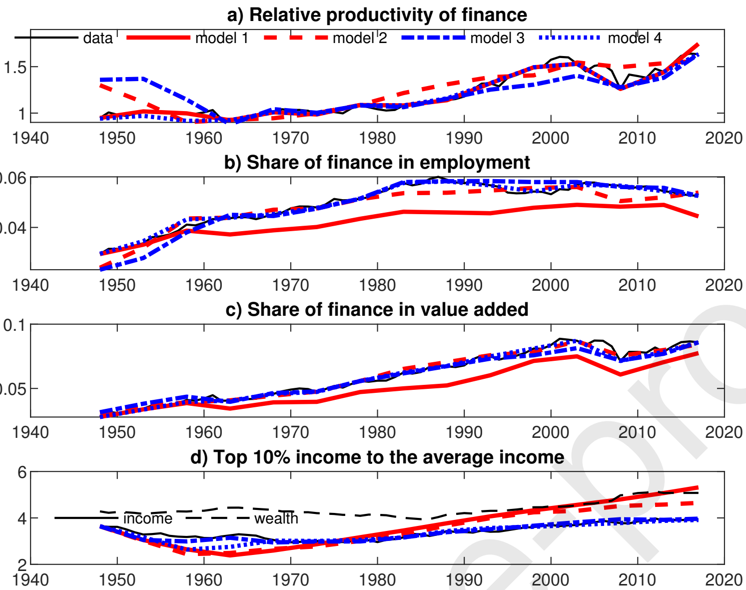
Figure 14 plots the performance of four calibrated models. The only source of time variation for Model 1 and Model 2 is the banker capacity, which is discussed in the details in the main text. Model 1 is the baseline calibration, which targets the relative productivity of finance. Model 2 targets jointly three times series: finance share in employment and value-added, and a measure of inequality. As for the case of Model 2, Model 3 and Model 4 target jointly three times series, but differently have an additional degree of freedoms on the top of time variation in M_t. While Model 3 adds the time variation in the return to scale on capital α_t^k, Model 4 has the time variation in the entrepreneurial bargaining power δ_t. Not surprisingly, an extra degree of freedom boosts the explanatory ability of the model. With time variation in two parameters, the model accounts for more than 90% of variation in all three time-series of interest, such as finance share in employment and value-added, and a measure of inequality.

Figure 15 plots the calibrated values of the capacity of bankers M_t, α_t^k return to scale on capital, δ_t the bargaining power of entrepreneurs, which determines the surplus split, and the relative productivity of talented agents in entrepreneurship $\left(\frac{z_t^H}{z_t^L}\right)^{\alpha_z}$. The banker capacity M_t increases over time in all models, which is reassuring that M_t is the main driver of the growth of finance. What is different is the speed of the banker capacity growth, which ranges from 5.3 % annually for the baseline model (Model 1) to 12 % annually for Model 3 and 4. The dynamics of entrepreneurial bargaining power δ_t might reflect the Venture capital boom prior to the burst of the dot-com bubble. The U-shape dynamics of return to scale on capital might reflect a dip in the markups from the early 1970s until the late 1980s (Diewert and Fox, 2008).

Appendix B.4 The Feedback from Inequality

In this part, we investigate the importance of feedback from inequality to the size of finance. Figure 16 compares Model 2, which targets three time-series of interest with only time variation in the banker capacity to Model 3 and Model 4. Model

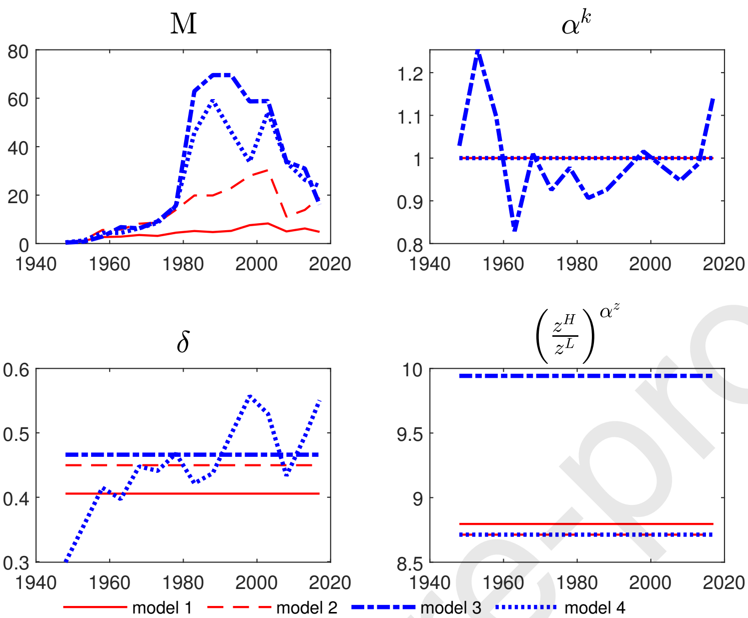
Figure 14. Targeted and untargeted series: extension



Note: The figure compares the performance of two models against four data series. Model 1 target the time series of the relative productivity of finance by varying in the banker capacity M_t. Model 2, Model 3 and Model 4 target jointly three times series: the financial shares in employment and value-added, and a measure of inequality, which is measured as the ratio of top 10% income to the average income. While Model 3 adds the time variation in the return to scale on capital α_t^k, Model 4 has the time variation in the entrepreneurial bargaining power δ_t The solid black line represents the data. The solid red line corresponds to Model 1 and the dashed red line corresponds to Model 2; the blue dash-dot line corresponds to Model 3, the blue dotted line corresponds to Model4. Panel a) reports the relative productivity of finance. Panel b) reports the share of finance in employment. Panel c) reports the share of finance in value-added. Panel d) plots the ratio of top 10% income to the average income. Additionally, the dashed black line represents the ratio of top 10% wealth to the average wealth in the last panel.

3 and Model 4 serve to quantify the importance of endogenous feedback inequity. Model 3, the blue dash-dot line, shares the same parameters as Model 2, but keeping the inequality constant at the level in the first period $\frac{k_t^H}{k_t^L} = \frac{k_0^H}{k_0^L}$. Model 4, the blue dotted line, is the same as Model 3, but I recalibrate the banker capacity M_t in order to match the size of the finance. In all four models, the change in the banker capacity is an important driver for the size of finance. By comparing Model 2 to Model 3 and 4, we can conclude that the endogenous feedback from the size of finance on inequality and vice versa is quantitatively important. Without this feedback, we either overstate the growth of finance in terms of value-added (Model 3) or understate the growth of finance in terms of employment (Model 4). This confirms the findings of the previous section, such as wealth inequity

Figure 15. Time varying parameters



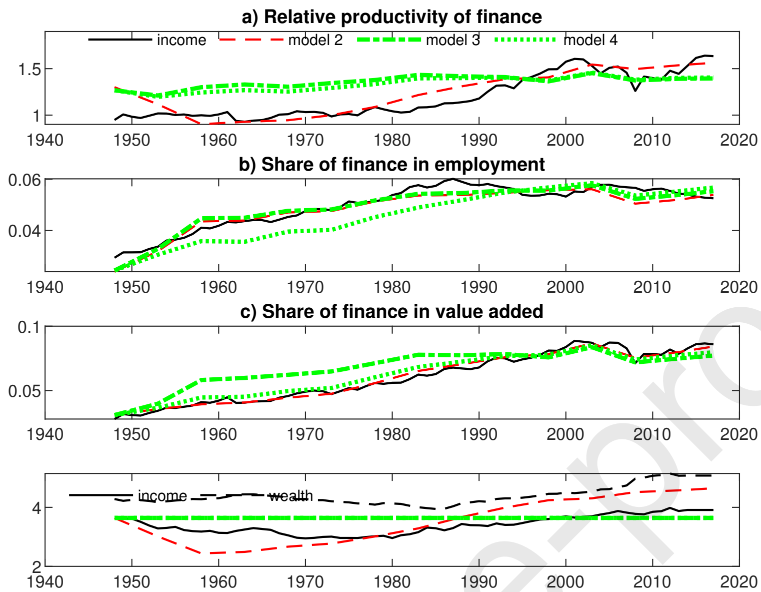
Note: This figures plot the dynamics of the banker capacity M for two calibrated models. The solid red line represents Model 1, which is the baseline calibration with the relative finance productivity is the sole target. The dashed red line represents Model 2. The targets of Model 2 are three times series: finance share in employment and value added, and a measure of inequality.

plays an important role in determining the equilibrium size of the financial sector. Furthermore, the resulting relative productivity of finance in Model 3 and Model 4 is completely counterfactual.

Appendix B.5 Does the model suit for the financial sector as a whole?

The financial sector provides many useful functions to the economy, as discussed in section 1.3. This paper focuses on two services: intermediation and sorting between investors and entrepreneurs. The model can reasonably apply to an intermediation activity, where there is asymmetric information; significant resources are spent on trying to overcome information asymmetry, and matching frictions are present. Information superiority of the financial sector with respect to ordinary investors (households) is a fairly standard assumption in finance literature supported by empirical evidence (Durnev et al., 2004). Even though many intermediation activities fall into this category (consumer credits, mortgages and corporate credits,

Figure 16. Targeted and untargeted series



Note: The figure compares the performance of two models against four data series. All three models exploit time variation in the banker capacity M_t and share the same targets, which are financial shares in employment and value-added, and a measure of inequality, which is measured as the ratio of top 10% income to the average income. While The solid black line represents the data. The dashed red line corresponds to Model 2. Model 3, the green dash-dot line, shares the same parameters as Model 2, but keeping the inequality constant at the level in the first period $\frac{k_t^H}{k_t^L} = \frac{k_0^H}{k_0^L}$. Model 4, the green dashed line, is the same as Model 3, but banker capacity M_t is recalibrated in order to match the size of finance. Panel a) reports the relative productivity of finance. Panel b) reports the share of finance in employment. Panel c) reports the share of finance in value-added. Panel d) plots the ratio of top 10% income to the average income. Additionally, the dashed black line represents the ratio of top 10% wealth to the average wealth in the last panel.

insurance), private equity finance and venture capital are subindustries for which the assumptions of the model are particularly valid.

It is important to know several facts about private equity finance and venture capital (VC). First, as shown by Greenwood and Scharfstein (2013), private equity finance and VC contributes substantially to the overall growth of the financial sector (almost a third of the total growth). Second, a private equity fund precisely does matching between a few selected young and fastly growing firms (talented entrepreneurs) and high-net-worth individuals (capital-abundant investors). The private equity fund provides an opportunity to invest in a few companies over a long-term horizon for a small number of wealthy investors. Third, it is not a competitive market: an entrepreneur receives only several offers from VC firms.

Gordon (2000) shows that 71.32% of firms reported they had received more than one offer to invest from VC firms. The mean number of offers was 3.18. This means entrepreneurs have a choice about which a VC firm to invest in their companies, but this choice is somewhat limited, and matching plays an important role. Furthermore, these offers are clearly not public information. Fourth, the restriction on the number, rather than the amount, reflects anecdotal evidence that VCs' scarce resource is time, not money, and that deals require roughly equal amounts of time. In a systematic study of VCs' investment analyses, Kaplan and Stromberg (2001) find that time commitment is a common concern for VCs when evaluating potential investments. These results confirm that VCs spend a great deal of time and effort in evaluating and screening transactions. This is consistent with anecdotal accounts that the scarcest commodity a VC has is time, not capital (Gladstone, 1988).

Appendix B.6 Investment banking and private equity

Most global banks, such as Credit Suisse, Barclays, BNP Paribas, Citibank, Deutsche Bank, HSBC, JPMorgan Chase, and UBS, have a separate business unit with dedicated teams of client advisors and product specialists exclusively for high-net-worth individuals. They provide a wide range of investment opportunities, including bonds, stocks, and, more importantly, private equity finance.

Private equity is an important channel through which long-term investments are made. It has grown steadily over the past three decades, and today private equity funds worldwide manage over $1 trillion. For some countries, such as Israel, the US, and the UK, private equity accounts for more than 5% of total investment (see Table 3 for details).

Table 3. The size of private equity

	% GDP		% Investment	
	2010	2011	2010	2011
Israel	0.63	2.09	3.50	10.45
UK	1.13	0.75	7.53	5.00
US	0.9	0.98	5.00	5.44
China	0.16		0.33	
World		0.30		1.58

I would like to convince the reader that the matching assumption definitely holds for private equity. A small private equity fund provides an opportunity to invest in a few companies over a long-term horizon for a small number of wealthy investors. As we can see from Table 4, private equity funds typically employ 9 professionals. These professionals select one or two companies each for the fund to invest in. Investments are large (over $50 million). Investors are wealthy and expected to invest over a long-term horizon. The minimum required commitment rises from a median of $1 million for funds of $100 million or less, up to a median of $10 million for funds of $1 billion or more. There is no active market for private equity positions, making these investments illiquid and difficult to value. Private equity funds typically have horizons of 10–13 years, during which the invested capital cannot be redeemed.

Table 4. Private equity funds

		Number of			Size ($ mn)	
	N	professionals	investments	I/P	Fund	Investment
VC	94	9	20	2	225	11.25
Buyout	144	13	12	1	600	50

Given the long-term horizon and the high entry costs, the question is why investors are willing to engage in these investments. Investors are compensated well by substantially higher returns. Table 5 shows that the return from an investment in private equity funds is three times higher than in stocks. We can see the comparison with inflation and the returns on other assets: stocks, gold, T-bills, etc.

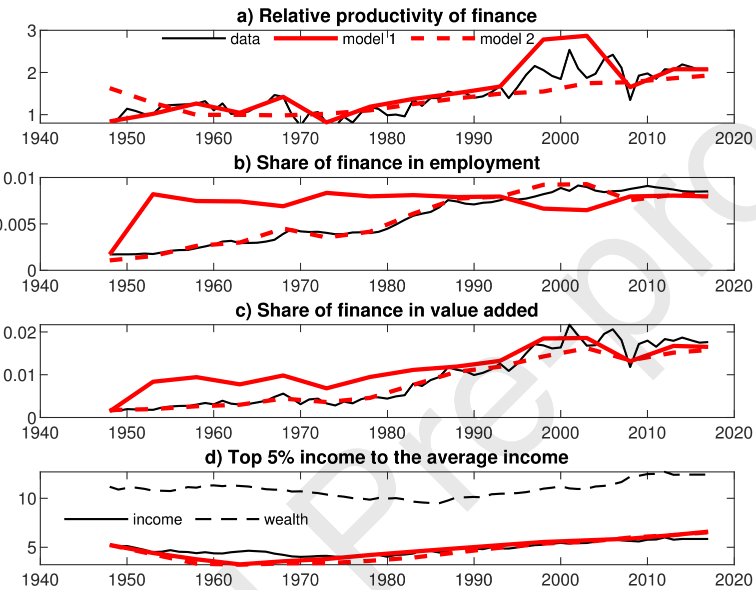
Table 5. US real asset returns

Period	PEF	S&P	TBond	Gold	Inflation
1997-2011	9.2%	3.2%	0.4%	7.0%	2.4%
1975-2011		7.5%	1.3%	4.0%	4.2%

Appendix B.7 New finance

Figure 17 shows the comparison between the data for "new finance", described in subsection Appendix B.1, and the outcomes of two calibrated models. The top panel presents the relative finance productivity. Panel b) and Pabel c) present the employment share of finance and the value-added share of finance accordingly. The bottom panel presents the ratio of top 5% income to mean income. The solid black line and the dashed black line are data.Model 1 is the baseline calibration, which targets the relative productivity of finance. Model 2 targets jointly three times series: finance share in employment and value added, and a measure of inequality. By comparing Models 1 and 2, we can conclude that it is enough to have only time-varying M_t in order to match simultaneously the evolution of the size of the financial sector in employment and value-added as well as inequality.

Figure 17. Targeted and untargeted series

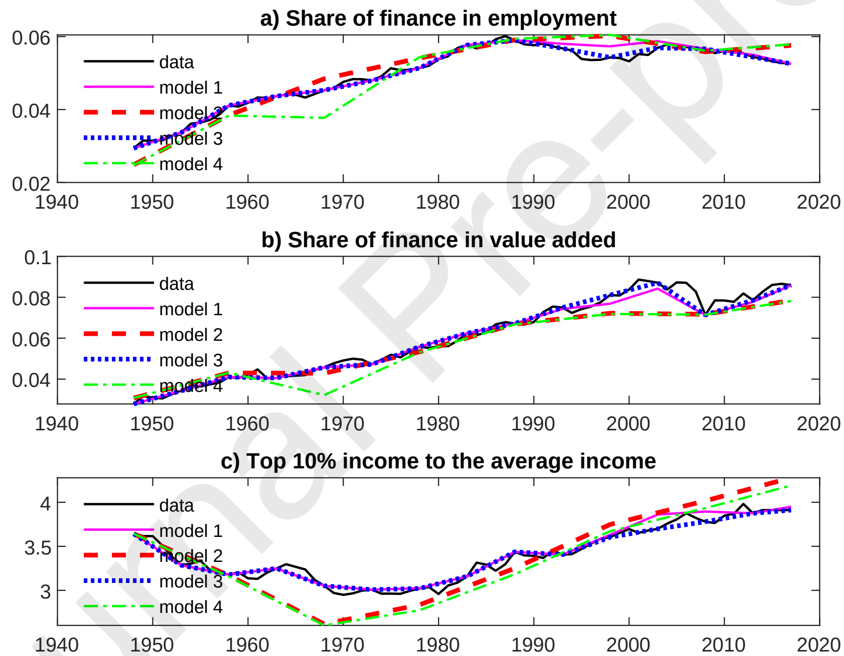


Note: The figure compares the performance of two models against four data series. While Model 1 is the baseline calibration, where the only target is the time series of the relative productivity of finance, for Model 2, the targets are three time series, including the financial shares in employment and value-added, and a measure of inequality, which is measured as the ratio of top 5% income to the average income. The solid black line represents the data. The solid red line corresponds to model 1 and the dashed red line corresponds to Model 2. Panel a) reports the relative productivity of finance. Panel b) reports the share of finance in employment. Panel c) reports the share of finance in value-added. Panel d) plots the ratio of top 5% income to the average income. Additionally, the dashed black line represents the ratio of top 5% wealth to the average wealth in the last panel.

Appendix B.8 The US tax system

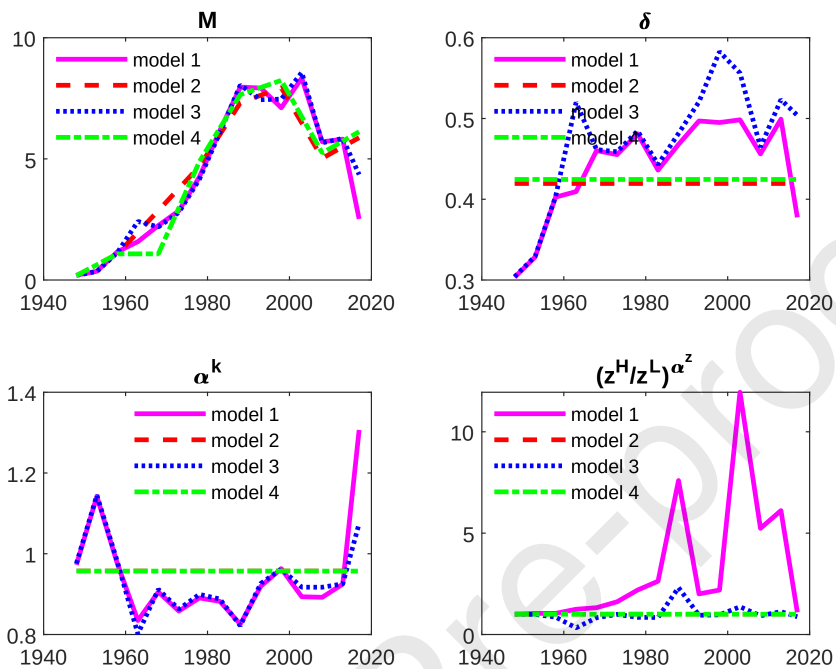
In these subsections, the impact of the changes in the US tax system is studied. As in the baseline calibration, the target is the post-tax top 10% average to the overall average. However, instead of using the average tax rates for the top 10 % income and for the average income, the actual tax rates from Figure 10 are feed into the model. Figure 18 shows the comparison between the data and the outcome of four calibrated models. The top panel presents the employment share of finance. The middle panel presents the value-added share of finance. The bottom panel presents the ratio of top 5% wealth to median wealth over time. I compare the data (the solid black line) with four calibrated models. Model 1 (the solid magenta line) and Model 2 (the dashed red line) are the models from the baseline calibration. Model 3 and Model 4 are results of the re-estimation of Model 1 and Model 2 based on the evolution of actual tax rates. As long as the targets are the same by comparing Model 3 to Model 1 and Model 4 to Model 2, we can conclude that feeding actual tax rates into the model does not alter significantly. The change of taxation leads to somewhat higher levels of bargaining power and the banker capacity and low levels of the relative productivity of talented agents in entrepreneurship.

Figure 18. Impact of Taxation: Models vs. Data



Notes: The figure contrasts the four estimated models to the data. The solid black line is the data. Model 1, the solid magenta line, is the most flexible model with four time-varying parameters (M,N, δ, α^k). Model 2, the dashed red line, represents a more restrictive case, where the two time-varying parameters are M and N. Model 3, the green dash-dot line, is the most restrictive case with one time-varying parameter M. The top panel plots the employment share of finance. The middle panel plots the value-added share of finance. The bottom panel plots inequality is measured as the ratio of the top 5% average income to the overall average. The parameters are estimated using simulation methods of moments.

Figure 19. Impact of Taxation: Estimated Parameters



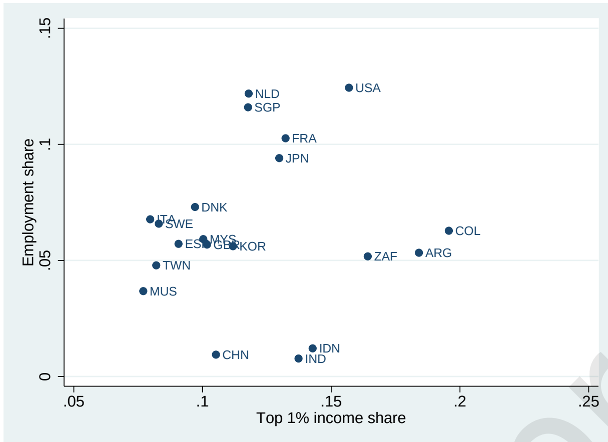
Notes: The figure plots four estimated parameters for four models: the banker capacity M (the number of investors which can be matched by a banker); δ the bargaining power of entrepreneurs, which determines the surplus split; the banker capacity N (the number of entrepreneurs which can be matched by a banker); α_k return to scale on capital.

Appendix B.9 Cross-country data

In this section, I provide cross-country evidence to answer the question of how the distribution of wealth and talent affects the size of the financial sector. As predicted by the model, the evidence clearly shows a positive relationship between the size of the financial sector and the inequality of wealth and talent. Even though the model predicts a causal link from the joint distribution of wealth and talent to the equilibrium size of the financial sector, in this section, I intend to make no causal statement. Let me start with a simple cross-section, Figure 20 plots the average share of employment in the financial sector versus the average top 1 % income share. As we can see, there is a strong positive relationship between the two. Interestingly, we obverse the clustering of developed and developing counties along two parallel lines.

To test the relationship, I need to have a compatible cross-country measure of moments of talent and wealth distributions, and the size of the financial sec-

Figure 20. The size of finance vs inequality



Notes: The figure plots plots the average share of employment in the financial sector versus the average top 1 % income share.

tor. Unfortunately, data availability limits the choice. For the talent distribution, I employ two proxies. The first one is the score in the Programme for International Student Assessment (PISA). The PISA test aims to evaluate education systems worldwide every three years by assessing 15-year-olds' competencies in key subjects: reading, mathematics, and science. To date, over 70 countries have participated in PISA. It is a widely used measure for cross-country comparisons of students' performance. The PISA data is available for the years 2003, 2006, and 2009. I choose the mean and variance of 2009 science scores in the PISA test as a proxy for the moments of talent distributions because it includes the highest number of countries. Moreover, this choice hardly affects the results, because PISA scores are highly correlated over time and disciplines: the correlation coefficients exceed 0.97. Second, there is extensive evidence that talent positively affects the obtained level of education. Furthermore, education is a good predictor of the success of entrepreneurial activity. I use the share of entrepreneurs with a college degree from the Global Entrepreneurship Monitor (GEM) 2001 - 2015 APS Global Key Indicators to proxy for the talent distribution. This proxy is less preferred than the first one because it is highly unlikely that this measure of talent suffers from reverse causality or the missing third factor. There is no reason why the size of the financial sector today might affect the performance of secondary school students today.

To the best of my knowledge, there is no cross-country data on wealth inequality. Therefore, I use income distribution as a proxy for wealth distribution. Income inequality is a fairly standard proxy for wealth inequality but may underestimate it. Income and wealth are not particularly well correlated either at the individual level for a given point. However, if we measure the correlation over time between top income and wealth shares for a particular country, for example, the US, we observe that the shares are highly correlated. The more concentrated are the shares, the higher are the correlations between them. Furthermore, income shares are more volatile and tend to lead wealth shares. Therefore, the ten years moving average of the top 1% income share is the preferred proxy for wealth inequality. We can see the comovement of wealth and income shares in the US in Appendix B.5.

The alternative measure of income inequality, which I employ, is the Gini indexes from the Standardized World Income Inequality Database (SWIID), Version 3.0, and top income shares from the World Wealth and Income Database (WID). The SWIID provides comparable Gini indexes of gross and net income inequality for 173 countries for as many years as possible from 1960. The WID includes 45 countries, for some going up to a century Solt (2016); Alvaredo et al. (2015). The last issue is how to measure the size of finance. I construct the share of financial industry employment in total employment using two datasets: the International Labour Organization dataset, which contains employment by economic activity for 165 countries starting from 1968, and the GGDC 10-Sector Database Version 2014, which contains industry-level data for employment and output for 31 countries from the 1960s up to the present (Timmer et al., 2015)

After conducting panel unit root tests, such as the Fisher combination test (Maddala and Wu, 1999) and the Pesaran (2007) panel unit root test, I calculate the growth rate of real GDP per capita to make it stationary. The specification of the full model is given by:

$$FI_{it} = \gamma_0 + \gamma_1 \Delta GDP_{it} + \gamma_2 II_{it} + \gamma_3 MT_i + \gamma_4 VT_i \tag{46}$$

where FI_{it} is the share of the financial sector; ΔGDP_{it} is the real GDP per capita growth; II_{it} is the proxy for income inequality; MT_i is the mean of talent distribution; VT_i is the variance of talent.

Table 6 reports the results of regressions for the share of finance in % of total employment. As you see, after controlling for the country fixed effect, the higher share of finance is positively associated with the higher inequity, which is robust to both proxies: Gini (columns (1), (2)) and top 1% income share(columns (5), (6)). As mentioned previously, measures of income inequality are more volatile than the measures of wealth inequality. If I use the ten-year moving average to smooth these fluctuations, the relations become stronger, as expected. Furthermore, the positive association with the inequity is mostly driven by high-income countries

(Compare columns (1) with (2) and (5) with (6)).

The data for talent distribution is cross-sectional. I replace the country's fixed effect with the proxies for talent distribution. Compare to columns (6) and (7), the coefficient of MA Top 1% remains unchanged. Furthermore, the variance of PISA scores is positively associated with the size of finance. However, this result is not very robust and is mostly driven by middle-income countries, which is consistent with the observation of Mayer-Haug et al. (2013) that entrepreneurial talent is more relevant in developing economies.

To summarize, the data suggests that the size of finance is positively and strongly associated with inequality, which is in line with the model. I provide some evidence for a weak relationship between the size of finance and talent inequality.

Table 6. The share of finance (in % of total employment) [10]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
GDP growth	-0.0911***	-0.480***	-0.0776*	-0.200***	-0.193***	-0.343***	-0.190***	-0.214***	-0.341***
	(0.0169)	(0.0525)	(0.0323)	(0.0582)	(0.0307)	(0.0522)	(0.0501)	(0.0540)	(0.0550)
GINI	0.000360	0.00166**							
	(0.000196)	(0.000634)							
MA GINI			0.00171***	0.00270***					
			(0.000439)	(0.000742)					
GDPPC0			0.0193***	0.0213			0.0390***	0.0175	0.00460
			(0.00580)	(0.0171)			(0.00925)	(0.0173)	(0.00973)
Mean PISA			0.000271**	0.000272			0.000884***	0.000301	
			(0.0000986)	(0.000495)			(0.000229)	(0.000566)	
Var. PISA			0.000676	0.00119			-0.00178	-0.00183	
			(0.000839)	(0.00162)			(0.00152)	(0.00178)	
Top 1%					0.167***				
					(0.0238)				
MA Top 1%						1.394***	1.045***	1.345***	1.294***
						(0.0974)	(0.0790)	(0.0876)	(0.0992)
The share of college									-0.260***
									(0.0725)
Effects	FE	FE, HI	MA	HI, MA	FE	FE, HI, MA	MA	HI, MA	HI, MA
N	1497	542	597	397	733	296	313	279	273
R2	0.0217	0.143	0.0311	0.0591	0.107	0.474	0.407	0.506	0.456

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: FE stands for country fixed effect; HI stands for high income countries according to OECD; MA stands for 10 years moving average. The employment data by industry is mostly from `www.ilo.org/ilostat` and supplemented by Timmer et al. (2015) and the national account data.

The results from Table 6 are consistent with the model. A higher share of financial employment is associated with more unequal income and talent distributions. The estimation results from the large sample, the last column with the Gini coefficient, show this clearly. Income inequality has an even stronger effect if I use the top income share instead of the Gini coefficient (columns 2 and 3). The result is not driven by country fixed effects (FE). We can see by comparing column 1 with FE and column 2 that the estimated coefficient of the top 5% share remains positive, significant, and almost unchanged.

REFERENCES

Acemoglu, Daron, 2002, Directed technical change, *The review of economic studies* 69, 781–809.

Acharya, Viral, Marco Pagano, and Paolo Volpin, 2016, Seeking alpha: Excess risk taking and competition for managerial talent, *The Review of Financial Studies* 29, 2565–2599.

Alvaredo, Facundo, Anthony B. Atkinson, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman, 2015, The world wealth and income database, accessed online: June, 30, 2015.

Arcand, Jean Louis, Enrico Berkes, and Ugo Panizza, 2015, Too much finance?, *Journal of Economic Growth* 20, 105–148.

Axelson, Ulf, and Philip Bond, 2015, Wall street occupations, *The Journal of Finance* 70, 1949–1996.

Axtell, Robert L, 2001, Zipf distribution of us firm sizes, *science* 293, 1818–1820.

Banerjee, Abhijit V, and Andrew F Newman, 1993, Occupational Choice and the Process of Development, *Journal of Political Economy* 101, 274–298.

Bauer, Christian, and Jose V. Rodriguez Mora, 2014, The joint determination of tfp and financial sector size, Working paper, mimeo.

Bazot, Guillaume, 2017, Financial consumption and the cost of finance: Measuring financial efficiency in europe (1950–2007), *Journal of the European Economic Association* jvx008.

Becker, Gary S, 1973, A Theory of Marriage: Part I, *Journal of Political Economy* 813–46.

Bernard, Andrew B, Jonathan Eaton, J Bradford Jensen, and Samuel Kortum, 2003, Plants and productivity in international trade, *American economic review* 93, 1268–1290.

Bienz, Carsten, and Julia Hirsch, 2011, The Dynamics of Venture Capital Contracts, *Review of Finance* 16, 157–195.

Böhm, Michael J, Daniel Metzger, and Per Strömberg, 2018, 'since you're so rich, you must be really smart': Talent and the finance wage premium, *Riksbank Research Paper Series* 18–3.

Bolton, Patrick, Tano Santos, and Jose A. Scheinkman, 2016, Cream-skimming in financial markets, *The Journal of Finance* 71, 709–736.

Bond, Philip, and Vincent Glode, 2014, The labor market for bankers and regulators, *Review of Financial Studies* .

Buera, Francisco J., and Joseph P. Kaboski, 2012, The rise of the service economy, *American Economic Review* 102, 2540–69.

Buera, Francisco J., and Yongseok Shin, 2013, Financial Frictions and the Persistence of History: A Quantitative Exploration, *Journal of Political Economy* 121, 221 – 272.

Cochrane, John H., 2013, Finance: Function Matters, Not Size, *Journal of Economic Perspectives* 27, 29–50.

Cooley, Thomas, Ramon Marimon, and Vincenzo Quadrini, 2020, Commitment in organizations and the competition for talent, *The Review of Economic Studies* 87, 2165–2204.

Célérier, Claire, and Boris Vallée, 2019, Returns to talent and the finance wage premium, *Review of Financial Studies* 32, 4005–4040.

Diewert, W Erwin, and Kevin J Fox, 2008, On the estimation of returns to scale, technical progress and monopolistic markups, *Journal of Econometrics* 145, 174–193.

Durnev, Art, Randall Morck, and Bernard Yeung, 2004, Value-Enhancing Capital Budgeting and Firm-specific Stock Return Variation, *Journal of Finance* 59, 65–105.

Galor, Oded, and Joseph Zeira, 1993, Income Distribution and Macroeconomics, *Review of Economic Studies* 60, 35–52.

Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny, 2014, Finance and the Preservation of Wealth, *The Quarterly Journal of Economics* 129, 1221–1254.

Gertler, Mark, and Nobuhiro Kiyotaki, 2010, Financial Intermediation and Credit Policy in Business Cycle Analysis, in *Handbook of Monetary Economics*, chapter 11, 547–599 (Elsevier).

Gladstone, David, 1988, *Venture Capital Handbook* (Englewood Cliffs, NJ: Prentice Hall).

Glode, Vincent, and Richard Lowery, 2015, Compensating financial experts, *The Journal of Finance* .

Goldin, Claudia, and Lawrence F. Katz, 2008, Transitions: Career and family life cycles of the educational elite, *American Economic Review* 98, 363–69.

Gompers, Paul, and Josh Lerner, 2000, Money chasing deals? The impact of fund inflows on private equity valuation, *Journal of Financial Economics* 55, 281–325.

Gordon, Platt, 2000, The americas, *Global Finance* 14.

Gourinchas, Pierre-Olivier, and Jonathan A. Parker, 2002, Consumption Over the Life Cycle, *Econometrica* 70, 47–89.

Greenwood, Jeremy, and Boyan Jovanovic, 1990, Financial development, growth, and the distribution of income, *Journal of Political Economy* 98, 1076–1107.

Greenwood, Robin, and David Scharfstein, 2013, The Growth of Finance, *Journal of Economic Perspectives* 27, 3–28.

Guvenen, Fatih, Burhanettin Kuruscu, and Serdar Ozkan, 2014, Taxation of human capital and wage inequality: A cross-country analysis, *Review of Economic Studies* 81, 818–850.

Hosios, Arthur J, 1990, On the Efficiency of Matching and Related Models of Search and Unemployment, *Review of Economic Studies* 57, 279–98.

Hsieh, Chang-Tai, and Peter J. Klenow, 2009, Misallocation and manufacturing tfp in china and india, *The Quarterly Journal of Economics* 124, 1403–1448.

Jovanovic, Boyan, and Balazs Szentes, 2013, On the Market for Venture Capital, *Journal of Political Economy* 121, 493 – 527.

Kaplan, Steven N, and Joshua Rauh, 2013, It's the market: The broad-based rise in the return to top talent, *Journal of Economic Perspectives* 27, 35–56.

Kaplan, Steven N., and Per Stromberg, 2001, Venture Capitals As Principals: Contracting, Screening, and Monitoring, *American Economic Review* 91, 426–430.

Kaplan, Steven N., and Per Stromberg, 2003, Financial contracting theory meets the real world: An empirical analysis of venture capital contracts, *Review of Economic Studies* 70, 281–315.

Karabarbounis, Loukas, and Brent Neiman, 2014, The global decline of the labor share, *The Quarterly journal of economics* 129, 61–103.

Kneer, Christiane, 2013, The Absorption of Talent into Finance: Evidence from U.S. Banking Deregulation, Working Papers 391, De Nederlandsche Bank.

Kuhn, Moritz, Moritz Schularick, and Ulrike I Steins, 2020, Income and wealth inequality in america, 1949–2016, *Journal of Political Economy* 128, 3469–3519.

Kurlat, Pablo, 2019, The social value of financial expertise, *American Economic Review* 109, 556–90.

Levine, Ross, 2005, Finance and growth: Theory and evidence, in Philippe Aghion, and Steven Durlauf, eds., *Handbook of Economic Growth*, chapter 12, 865–934 (Elsevier).

Lucas, Robert Jr., 1988, On the mechanics of economic development, *Journal of Monetary Economics* 22, 3–42.

Lucas, Robert Jr., and Leonard Rapping, 1969, Real Wages, Employment, and Inflation, *Journal of Political Economy* 77, 721–54.

Maddala, G S, and Shaowen Wu, 1999, A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test, *Oxford Bulletin of Economics and Statistics* 61, 631–52.

Mayer-Haug, Katrin, Stuart Read, Jan Brinckmann, Nicholas Dew, and Dietmar Grichnik, 2013, Entrepreneurial talent and venture performance: A meta-analytic investigation of SMEs, *Research Policy* 1–23.

Metrick, Andrew, and Ayako Yasuda, 2010, The economics of private equity funds, *The Review of Financial Studies* 23, 2303–2341.

Midrigan, Virgiliu, and Daniel Yi Xu, 2014, Finance and Misallocation: Evidence from Plant-Level Data, *American Economic Review* 104, 422–58.

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny, 1991, The Allocation of Talent : Implications for Growth, *The Quarterly Journal of Economics* 106, 503–530.

Nash, John, 1950, The bargaining problem, *Econometrica* 18, 155–162.

Nash, John, 1953, Two-person cooperative games, *Econometrica* 21, 128–140.

Pesaran, M. Hashem, 2007, A simple panel unit root test in the presence of cross-section dependence, *Journal of Applied Econometrics* 22, 265–312.

Philippon, Thomas, 2007, Why has the us financial sector grown so much? the role of corporate finance.

Philippon, Thomas, 2010, Financiers versus Engineers: Should the Financial Sector be Taxed or Subsidized?, *American Economic Journal: Macroeconomics* 2, 158–182.

Philippon, Thomas, 2015, Has the us finance industry become less efficient? on the theory and measurement of financial intermediation, *American Economic Review* 105, 1408–38.

Philippon, Thomas, and Ariell Reshef, 2012, Wages and Human Capital in the U.S. Finance Industry: 1909–2006, *The Quarterly Journal of Economics* 127, 1551–1609.

Piketty, Thomas, 2014, *Capital in the twenty-first century* (The Belknap Press of Harvard Univ. Press, Cambridge, Mass. [u.a.]).

Piketty, Thomas, and Gabriel Zucman, 2014, Capital is back: Wealth-income ratios in rich countries, 1700-2010, *The Quarterly Journal of Economics* .

Restuccia, Diego, and Richard Rogerson, 2008, Policy distortions and aggregate productivity with heterogeneous plants, *Review of Economic Dynamics* 11, 707–720.

Romer, Paul M, 1986, Increasing Returns and Long-Run Growth, *The Journal of Political economy* 94, 1002–1037.

Rosen, Sherwin, 1981, The economics of superstars, *The American economic review* 71, 845–858.

Santarelli, Enrico, Luuk Klomp, and A.Roy Thurik, 2006, Gibrat's law: An overview of the empirical literature, in Enrico Santarelli, ed., *Entrepreneurship, Growth, and Innovation*, volume 12 of *International Studies in Entrepreneurship*, 41–73 (Springer US).

Scheidel, Walter, 2017, *The great leveler: violence and the history of inequality from the Stone Age to the twenty-first century* (Princeton University Press).

Shimer, Robert, and Lones Smith, 2001, Matching, search, and heterogeneity, *The B.E. Journal of Macroeconomics* 1, 1–18.

Shu, Pian, 2015, Innovating in Science and Engineering or 'Cashing In' on Wall Street? Evidence on Elite STEM Talent, Working papers, HBS.

Solt, Frederick, 2016, The standardized world income inequality database, *Social Science Quarterly* .

Timmer, Marcel, Gaaitzen de Vries, and Klaas de Vries, 2015, *Patterns of structural Change in Developing Countries*, 65–83, Routledge Handbooks (Routledge).

Wadhwa, Vivek, Ben A. Rissing, and G. Gereffi, 2006, Industry trends in engineering offshoring, report, Duke University.

Watanabe, Makoto, 2010, A model of merchants, *Journal of Economic Theory* 145, 1865–1889.

Wolff, Edward N, 2016, Household wealth trends in the united states, 1962 to 2013: What happened over the great recession?, *RSF: The Russell Sage Foundation Journal of the Social Sciences* 2, 24–43.