Cream skimming in financial markets*

Patrick Bolton
Columbia University and NBER

Tano Santos
Columbia University and NBER

Jose A. Scheinkman
Princeton University and NBER

February 25, 2011

Abstract

We propose a model where agents choose to become entrepreneurs or informed dealers in financial markets. Agents incur costs to become dealers and develop skills for valuing assets. The financial sector comprises a transparent exchange, where uninformed agents trade, and an opaque over-the-counter (OTC) market, where dealers offer attractive terms for the best assets. Dealers provide incentives for entrepreneurs to originate good assets, but the opaqueness of the OTC market allows dealers to extract rents. By siphoning out good assets, the OTC market lowers the quality of assets in the exchange. In equilibrium, dealers' rents are excessive and attract too much talent to Finance.

*This article was previously circulated under the title "Is the Financial Sector too Big?" We thank seminar participants at the London Business School, London School of Economics, Rutgers Business School, MIT-Sloan, Bank of Spain, CEMFI, University of Virginia-McIntire, Washington University-Olin Business School, and the 6th Banco de Portugal Conference on Monetary Economics for their feedback and suggestions. We also thank Guido Lorenzoni for detailed comments.

What does the financial industry add to the real economy? What is the optimal organization of financial markets, and how much talent is required in the financial industry? We revisit these fundamental questions in light of recent events and criticisms of the financial industry. Most notably, the former chairman of the Federal Reserve Board, Paul Volcker, recently asked:

> How do I respond to a congressman who asks if the financial sector in the United States is so important that it generates 40% of all the profits in the country, 40%, after all of the bonuses and pay? Is it really a true reflection of the financial sector that it rose from $2\frac{1}{2}\%$ of value added according to GNP numbers to $6\frac{1}{2}\%$ in the last decade all of a sudden? Is that a reflection of all your financial innovation, or is it just a reflection of how much you pay? What about the effect of incentives on all our best young talent, particularly of a numerical kind, in the United States?[1] [*Wall Street Journal*, December 14, 2009]

The core issue underlying these questions is whether the financial industry extracts excessively *high rents* from the provision of financial services, and whether these rents attract too much *young talent*.[2] In this paper we propose a new model of the financial sector which is segmented into two broad types of markets: on the one hand, there are organized, regulated, standardized, and transparent markets, where most retail ('*plain vanilla*') transactions take place, and on the other, there are informal, opaque, markets where *informed* (talented) dealers provide '*bespoke*' services to their clients. The *opaqueness* of prices and conditions of trade characterize many, though certainly not all, over-the counter (OTC) markets and for brevity we will refer to the set of opaque markets as OTC markets and to the transparent, standardized, markets as organized exchanges. We argue that, while OTC markets provide

[1]As Philippon (2008) points out this growth of the US financial industry is not just a result of globalization of financial services as it starts before globalization. Moreover, the US is *not* a significant net exporter of financial services.

[2]Goldin and Katz (2008) document that the percentage of male Harvard graduates with positions in Finance 15 years after graduation tripled from the 1970 cohort to the 1990 cohort, largely at the expense of occupations in law and medicine. We elaborate on this observation in Section 6 below.

indispensable valuation services to their clients, their opacity also allows informed dealers to extract too high rents. What is more, OTC markets tend to undermine organized exchanges by "*cream-skimming*" the juiciest deals away from them.

Our model also allows for an endogenous occupational choice between a financial and a real sector. We show that the excessively high informational rents obtained by informed dealers in the OTC markets tend to attract too much talent to the financial industry.

In our model, secondary market trading requires information about underlying asset quality and valuation skills. When an entrepreneur is looking to sell assets or his firm in the secondary market, the buyer must be able to determine the value of the assets that are up for sale. This is where the *talent* employed in the financial industry manifests itself. *Informed dealers* in the OTC market are better able to determine the value of assets for sale and can *cream-skim* the most valuable assets. Importantly, by identifying the most valuable assets and by offering more attractive terms for those assets than are available in the organized market, informed dealers in the OTC market also provide incentives to entrepreneurs to originate good assets. However, the central efficiency question for agents' occupational choices between the financial and real sectors is what share of the incremental value of these good assets dealers get to appropriate. Valuation skills in reality and in our model are costly to acquire and generally scarce. This is why not all asset sales can take place in the OTC market. Those assets that cannot be absorbed by the OTC market end up on the organized exchange. The relative scarcity of *informed capital* in the OTC market is a key determinant of the size of the information rents that are extracted by the financial sector in equilibrium.

The OTC market is an informal market where sellers of assets match with informed dealers and negotiate terms bilaterally. In this market price offers of dealers and negotiated transactions are not disclosed. This is in contrast to organized markets, where all quotes and transactions are posted. As a result of the scarcity of informed dealers and the *opacity* of the OTC market, informed dealers are able to extract an *informational rent* from the entrepreneurs selling the most valuable assets to them. Entrepreneurs with good assets can either sell their asset in the organized market, where it gets pooled with all other assets and therefore will be undervalued, or they can negotiate a better price with an informed dealer in the OTC market.

The cream-skimming activity of informed dealers imposes a negative price externality on the organized market, as uninformed investors operating in this market understand that they only get to buy an adversely selected pool of assets. This negative price externality in turn weakens the bargaining position of entrepreneurs selling good assets in the OTC market, as their threat-point of selling the asset in the organized market becomes less attractive.[3]

This negative externality on organized markets is why our model generates a counter-intuitive comparative statics result: rent-extraction by informed dealers actually increases as the OTC market expands. This is also the reason why: *i)* there is excessive information rent-extraction in the OTC market; and consequently, *ii)* the financial OTC market attracts too much talent.

In our model human capital invested to become an informed dealer serves as much to extract informational rents as to create social surplus (by incentivizing entrepreneurs to originate good assets). We show that in any equilibrium in which there is a sufficient number of informed dealers to provide incentives to entrepreneurs to originate good assets, there are in fact too many informed dealers relative to the social optimum. Our model thus helps explain how excessive rent extraction and entry into the financial industry can be an *equilibrium outcome*, and why competition for rents doesn't eliminate excessive rent extraction.

The structure of the financial industry, combining an opaque OTC market and an organized exchange is a key feature of our theory. Unlike models of informed trading in the tradition of Grossman and Stiglitz (1980), in our model dealer information in the OTC market is asset specific and is not reflected in a market price, as each transaction is an undisclosed bilateral deal between the dealer and the seller. Therefore, when more dealers compete in

[3]This form of cream-skimming is different from the cream-skimming considered by Rothschild and Stiglitz (1976) in insurance markets with adverse selection. In the insurance setting, insurers are uninformed about risk types, but offer contracts that induce informed agents (in particular, low risk types) to self-select into insurance contracts. For an application of the Rothschild-Stiglitz framework to competition among organized exchanges see Santos and Scheinkman (2001).

the OTC market, this does not result in more information being transmitted. On the contrary, more competition by informed dealers results in more cream-skimming and more information rent extraction. As a result information is overproduced instead of underproduced. Our highly stylized model can be seen as an *allegory* of a general phenomenon in the financial industry: informed parties have an incentive to trade and remove themselves from organized markets.

Indeed, this phenomenon is not just present in corporate equity and bond markets. It is also a feature of derivatives, futures, and swaps markets.The coexistence of OTC forwards and futures contracts traded on exchanges provides an interesting illustration. Why don't all future transactions take place on organized futures markets? One reason is as in our model: transactions in forward markets are primarily between informed dealers and producers who seek to hedge against spot-price movements. By trading in forward markets these producers are typically subject to lower margin calls when the spot price moves away from the forward price. The reason is that informed dealers understand that (as long as they are not over-hedged) producers actually benefit from movements in spot price away from the forward price and therefore do not give rise to higher counterparty risk. As a result, a substantial portion of commodities production is hedged outside exchanges, via forward contracts with banks and trading companies. These contracts give producers less favorable prices, but require smaller margins. After doing due diligence to verify that a producer is not over-hedged, a bank can feel confident that it will actually be better off if spot prices increase. This same bank would most likely also engage in an opposite forward with a counterparty for whom buying forwards would actually lower risk, and only hedge the net amount with futures contracts. Thus, by demanding a uniform mark-to-market margin of all parties exchanges induce a lower mix of producer-hedgers, and hence a riskier set of buyers and sellers.

Figure 1, from Philippon and Resheff (2008), plots the evolution of US wages (relative to average non-farm wages) for three subsegments of the finance services industry: credit, insurance and 'other finance.' Credit refers to banks, savings and loans and other similar institutions, insurance to life and P & C, and 'other finance' refers to the financial investment industry and investment banks. As the plot shows the bulk of the growth in remuneration in the financial industry took place in 'other finance.' An increasing fraction of activities in 'other finance' takes place in opaque markets where information is particularly valuable and our model argues that this opacity allows informed dealers to extract excessively high rents.

The paper is organized as follows: Section 2 outlines the model. Section 3 analyzes entrepreneurs' moral hazard in origination problem and describes some basic attributes of equilibrium outcomes. The analysis of welfare and equilibrium allocation of talent in financial markets is undertaken in section 4. Section 5, in turn, considers the robustness of our main results to the situation where informed dealers compete with each other. Finally, Section 6 offers further discussion on the model as well as examples and applications. Section 7 concludes. All proofs are in the Appendix.

Related Literature. In his survey of the literature on financial development and growth, Levine (2005) synthesizes existing theories of the role of the financial industry into five broad functions: 1) information production about investment opportunities and allocation of capital; 2) mobilization and pooling of household savings; 3) monitoring of investments and performance; 4) financing of trade and consumption; 5) provision of liquidity, facilitation of secondary market trading, diversification, and risk management. As he highlights, most of the models of the financial industry focus on the first three functions, and if anything, conclude that from a social efficiency standpoint the financial sector is too small: due to asymmetries of information, and incentive or contract enforceability constraints, there is underinvestment in equilibrium and *financial underdevelopment*.

In contrast to this literature, our model mainly emphasizes the fifth function in Levine's list: secondary market trading and liquidity provision. In addition, where the finance and growth literature only distinguishes between bank-based and market-based systems (e.g. Allen and Gale, 2000), a key departure of our model is the distinction we draw between markets in which trading occurs at prices and conditions that are not observable by other participants, and markets in which trading occurs under prices and conditions that are observed by all potential participants.[4]

Our paper contributes to a small literature on the optimal allocation of talent to the financial industry. An early theory by Murphy, Shleifer and Vishny (1991) (see also Baumol, 1990) builds on the idea of increasing returns to ability and rent seeking in a two-sector model to show that there may be inefficient equilibrium occupational outcomes, where too much talent enters one market since the marginal private returns from talent could exceed the social returns. More recently, Philippon (2008) has proposed an occupational choice model where agents can choose to become workers, financiers or entrepreneurs. The latter originate projects which have a higher social than private value, and need to obtain funding from financiers. In general, as social and private returns from investment diverge it is optimal in his model to subsidize entrepreneurship. Neither the Murphy et al. (1991) nor the Philippon (2008) models distinguish between organized exchanges and OTC markets in the financial sector, nor do they allow for excessive informational rent extraction through cream-skimming. In independent work Glode, Green and Lowery (2010) also model the idea of excessive investment in information as a way of strengthening a party's bargaining power. However, Glode et al. (2010) do not consider the occupational choice question of whether too much young talent is attracted towards the financial industry. Finally, our paper relates to the small but burgeoning literature on OTC markets, which, to a large extent, has focused on the issue of financial intermediation in the context of search models.[5] These papers have some common elements to ours, in particular the emphasis on bilateral bargaining when thinking about OTC markets, but their focus is on the liquidity of these markets and they do not address issues of cream-skimming or occupational choice.

[4]The literature comparing bank-based and market-based financial systems argues that bank-based systems can offer superior forms of risk sharing, but that they are undermined by competition from securities markets (see Jacklin, 1987, Diamond, 1997, and Fecht, 2004). This literature does not explore the issue of misallocation of talent to the financial sector, whether bank-based or market-based.

[5]See Duffie, Garleanu and Pedersen (2005), Vayanos and Wang (2007), Vayanos and Weill (2008), Lagos and Rocheteau (2009), Lagos, Rocheteau and Weill (2010) and Afonso (2010).

1 The model

We consider a competitive economy divided into two sectors–a real, productive, sector and a financial sector–and three periods $t = 0, 1, 2$.

1.1 Agents

There is a continuum of risk-neutral, agents who can be of two different types. Type 1 agents, of which there is a large measure, are *uninformed rentiers*, who start out in period 0 with a given endowment ω (their savings), which they consume in either period 1 or 2. Their preferences are represented by the utility function

$$u(c_1, c_2) = c_1 + c_2, \tag{1}$$

Type 2 agents form the *active population*. Each type 2 agent can choose to consume their endowment or work either as a (self-employed) *entrepreneur* in the real sector, or as a *dealer* in the financial sector. Type 2 agents make an occupational choice decision in period 0. Our parametric assumptions will insure that in equilibrium all type 2 agents choose to work.

We simplify the model by assuming that type 2 agents can only differ in their ability to become well-informed dealers. Specifically, we represent the mass of type 2 agents by the unit interval $[0, 1]$ and order these agents $d \in [0, 1]$ in increasing order of the costs they face of acquiring the human capital to become well informed dealers: $\varphi(d)$. That is, we assume that $\varphi(d)$ is non-decreasing. This assumption will imply that if an agent of type $\hat{d}$ prefers to become a dealer, so will all agents with $d \in [0, \hat{d})$. In addition we assume that there exists a $\overline{d} < 1$ such that for $d \geq \overline{d}$

$$\varphi(d) = +\infty. \tag{2}$$

Hence agents $d \geq \bar{d}$ always stay in the real sector.

In all other respects, type 2 agents are identical: They face the same *i.i.d.* liquidity shocks: they value consumption only in period 1 with probability π and only in period 2 with probability $(1 - \pi)$.[6] Their preferences are represented by the utility function

$$U(c_1, c_2) = \delta_1 c_1 + (1 - \delta)c_2, \tag{3}$$

where $\delta \in \{0, 1\}$ is an indicator variable and prob $(\delta = 1) = \pi$.

All type two agents have a unit of endowment in period 0. If a type 2 agent chooses to work in the real sector as an entrepreneur, he invests his unit endowment in a project in period 0. He then manages the project more or less well by choosing a hidden action $a \in \{a_l, a_h\}$ at *private effort cost* $\psi(a)$, where $0 < a_l < a_h \leq 1$. If he chooses $a = a_l$ then his *effort cost* $\psi(a_l)$ is normalized to zero, but he is then only able to generate a high output $\gamma\rho$ with probability a_l (and a low output ρ with probability $(1 - a_l)$), where $\rho \geq 1$ and $\gamma > 1$. If he chooses the high effort $a = a_h$, then his effort cost is $\psi(a_h) = \psi > 0$, but he then generates a high output $\gamma\rho$ with probability a_h. We assume, of course, that it is efficient for an entrepreneur to choose effort a_h:

$$(\gamma - 1)\rho\Delta a > \psi \qquad \text{where} \qquad \Delta a = a_h - a_l.$$

The output of the project is obtained only in period 2. Thus, if the entrepreneur learns that he wants to consume in period 1 ($\delta = 1$) he needs to sell claims to the output of his project in a financial market to either *patient* dealers, who are happy to consume in period 2, or rentiers, who are indifferent as to when they consume. For simplicity, we assume that in period 1 entrepreneurs have no information, except for the effort they applied, concerning the eventual output of their project. Note also that patient entrepreneurs have no output in period 1 that they could trade with impatient entrepreneurs.

If type 2 agent d chooses to work in the financial sector as a dealer, he saves his unit endowment to period 1, but incurs a utility cost $\varphi(d)$ to build up human capital in period 0. This human capital gives agent d the skills to value assets originated by entrepreneurs and that are up for sale in period 1. Specifically, we assume that a dealer is able to perfectly ascertain the output of any asset in period 2, so that dealers are *perfectly informed*. If dealers learn that they are patient ($\delta = 0$) they use their endowment, together with any collateralized borrowing,

[6] Our main results are robust to assuming that the liquidity shocks depend on the activity.

to purchase assets for sale by impatient entrepreneurs.[7] If they learn that they are impatient they simply consume their unit endowment. For simplicity, we assume that patient dealers can only acquire one unit of the asset at date 1.[8]

1.2 Financial Markets

An innovation of our model is to allow for a *dual financial system*, in which assets can be traded either in an over-the-counter (OTC) dealer market or in an organized exchange. Information about asset values resides in the OTC market, where informed dealers negotiate asset sales on a bilateral basis with entrepreneurs. On the organized exchange assets are only traded between uninformed rentiers and entrepreneurs. We also allow for a debt market where borrowing and lending in the form of *default-free* collateralized loans can take place. In this market a loan can be secured against an entrepreneur's asset. Since the lowest value of this asset is ρ, the default-free loan can be at most equal to ρ.

Thus, in period 1 an impatient entrepreneur has several options: i) he can borrow against his asset; ii) he can sell his asset for the competitive equilibrium price p in the organized exchange; iii) he can go to a dealer in the OTC market and negotiate a sale for a price p^d.

Consider first the OTC market. This market is composed of a measure $d(1-\pi)$ of *patient* dealers ready to buy assets from the mass $(1-d)\pi$ of *impatient* entrepreneurs. Each of the dealers is able to trade a total output of at most $1+\rho$, his endowment plus a maximum collateralized loan from rentiers of ρ, in exchange for claims on entrepreneurs' output in period 2. Impatient entrepreneurs turn to dealers for their information: they are the only agents that are able to tell whether the entrepreneur's asset is worth $\gamma\rho$ or just ρ. Just as in Grossman and Stiglitz (1980), dealers' information must be in scarce supply in equilibrium, as dealers must be compensated for their cost $\varphi(d)$ of acquiring their valuation skills. As will become clear below, this means not only that dealers only purchase high quality assets worth $\gamma\rho$ in

[7] By assuming that informed dealers know precisely the quality of the projects and entrepreneurs only know the effort they applied we are simplifying the asymmetric information problem.

[8] This can be justified by assuming that searching and managing assets demands the dealers time.

equilibrium, but also that not all entrepreneurs with high quality assets will be able to sell to a dealer.

In period 1 a dominant strategy for impatient entrepreneurs is to attempt to first approach a dealer. They understand that with probability $a \in \{a_l, a_h\}$ the underlying value of their asset is high, in which case they are able to negotiate a sale with a dealer at price $p^d > p$ with probability $m \in [0, 1]$. If they are not able to sell their asset for price p^d to a dealer, entrepreneurs can turn to the organized market in which they can sell their asset for p.

We show that in equilibrium only patient dealers and impatient entrepreneurs trade in the OTC market. We thus assume that the probability m is simply given by the ratio of the total mass of patient dealers $d(1 - \pi)$ to the total mass of high quality assets up for sale by impatient entrepreneurs, which in a symmetric equilibrium where all entrepreneurs choose the same effort level a is given by $a(1 - d)\pi$, so that

$$m(a, d) = \frac{d(1 - \pi)}{a(1 - d)\pi}. \tag{4}$$

Note that $m(a, d) < 1$ as long as d is sufficiently small and π is sufficiently large.[9] The idea behind this assumption is, first that any individual dealer is only able to manage one project at a time, and/or to muster enough financing to buy only one high quality asset. Second, in a symmetric equilibrium the probability of a sale of an asset to a dealer is then naturally given by the proportion of patient dealers to high quality assets.

The price p^d at which a sale is negotiated between a dealer and an entrepreneur is the outcome of bargaining (under symmetric information). The price p^d has to exceed the *status-quo* price p in the organized market at which the entrepreneur can always sell his asset. Similarly, the dealer cannot be worse off than under no trade, when his payoff is 1, so that the price cannot be greater than the value of the asset $\gamma\rho$. We take the solution to this bargaining game to be given by the *Asymmetric Nash Bargaining Solution*,[10] where the dealer has bargaining

[9]The assumption is formally made in expression (6) below.

[10]For a similar approach to modeling negotiations in OTC markets between dealers and clients see Lagos, Rocheteau, and Weill (2010).

power $(1-\kappa)$ and the entrepreneur has bargaining power κ (see Nash, 1950, 1953).[11] That is, the price p^d is given by

$$p^d = \arg \max_{s \in [p, \gamma\rho]} \{(s-p)^{\kappa}(\gamma\rho - s)^{(1-\kappa)}\},$$

or

$$p^d = \kappa\gamma\rho + (1-\kappa)p.$$

In a more explicit, non-cooperative bargaining game, with alternating offers between the dealer and entrepreneur à la Rubinstein (1982), the bargaining strength κ of the entrepreneur can be thought of as arising from a small probability per round of offers that the entrepreneur is hit by an *immediacy shock* and needs to trade immediately (before hearing back from the dealer) by selling his asset in the organized market. In that case the dealer would miss out on a valuable trade. To avoid this outcome the dealer would then be prepared to make a price concession to get the entrepreneur to agree to trade before this immediacy shock occurs (see Binmore, Rubinstein and Wolinsky, 1986).[12]

The price p^d may be higher than the dealer's endowment. In that case the dealer needs to borrow the difference $(p^d - 1)$ against the asset to be acquired. As long as this difference does not exceed ρ, the dealer will not be financially constrained. For simplicity, we restrict attention to parameter values for which the dealer is not financially constrained. We provide a condition below that ensures that this is the case.[13]

Consider next the organized exchange. We show that in equilibrium all assets of impatient entrepreneurs that are not sold in the OTC market trade. That is, $(1-a)(1-d)\pi$ low

[11] In Section 5 we show that our results are robust to assuming that the bargaining power of dealers decreases with the number of dealers.

[12] Symmetrically, there may also be a small *immediacy* shock affecting the dealer, so that the entrepreneur also wants to make concessions in negotiating an asset sale. Indeed, when a dealer is hit by such a shock the matched entrepreneur is unlikely to be able to find another dealer. More precisely, if θ is the probability per unit time that an entrepreneur or dealer is hit by an immediacy shock, and if α denotes the probability of an entrepreneur subsequently matching with another informed dealer then Binmore, Rubinstein and Wolinsky show that $\kappa = \alpha$.

[13] Note that the possibility that the dealer may be financially constrained may be another source of bargaining strength for the dealer. Exploring this idea, however, is beyond the scope of this paper.

quality assets and $(1-m)a(1-d)\pi$ high quality assets are sold in the exchange. The buyers of assets are uninformed rentiers, who are unable to distinguish high quality from low quality assets. Entrepreneurs also do not know the true underlying quality of their assets. A high quality asset pays $\gamma\rho - p$ and a low quality asset pays $\rho - p$. Thus the expected value of the assets traded in the exchange. is:

$$\frac{a(1-m)\gamma\rho + (1-a)\rho}{a(1-m) + (1-a)},$$

so that the competitive equilibrium price in the organized exchange is given by

$$p\,(a,d) = \frac{a(1-m)\gamma\rho + (1-a)\rho}{a(1-m) + (1-a)} = \frac{\rho[a(1-m)\gamma + (1-a)]}{1-am}, \tag{5}$$

where we have omitted the dependence of m on a and d, as in (4), for simplicity. Note also that p is decreasing in m, from the highest price $p = \rho[a(\gamma-1)+1]$ when $m = 0$ to the lowest price $p = \rho$ when $m = 1$.

1.3 Discussion and parameter restrictions

Our model of the interaction between the real and financial sector emphasizes the liquidity provision and valuation roles of the financial industry. It downplays the financing role of real investments. This role, which is emphasized in other work (e.g. Bernanke and Gertler, 1989 and Holmstrom and Tirole, 1997) can be added, by letting entrepreneurs borrow from either rentiers or dealers at date 0. The assets entrepreneurs sell in period 1 would then be net of any liabilities incurred at date 0. Since the external financing of real investments in period 0 does not add any novel economic effects in our model we have suppressed it.

The key interaction between the financial and real sectors in our model is in the incentives provided to entrepreneurs to choose high effort a_h when dealers are able to identify high quality assets and offer to pay more for these assets than entrepreneurs are able to get in the organized market. The social value of dealer information lies here. If it were not for these positive incentive effects, informed dealers would enrich themselves thanks to their cream-skimming activities in OTC markets but they would not create any net social surplus.

We have introduced ex-ante heterogeneity among type 2 agents only in the form of different utility costs in acquiring information to become a dealer. We could also have introduced heterogeneity in the costs of becoming an entrepreneur. We would then simply order type 2 agents in their increasing *comparative advantage* of becoming dealers and proceed with the analysis as in our current model. For simplicity we have therefore suppressed this added form of heterogeneity.

As we have argued above, we shall restrict attention to parameter values for which the measure of patient dealers is smaller than the measure of high quality assets put on the market by impatient entrepreneurs in period 1, so that

$$m(a, \overline{d}) = \frac{\overline{d}(1-\pi)}{a(1-\overline{d})\pi} < 1 \qquad \text{for} \qquad a \in \{a_l, a_h\}, \tag{6}$$

where, recall, $\overline{d}$ is defined in expression (2). Under this assumption dealers are always on the short side in the OTC market, which is partly why they are able to extract informational rents. Although it is possible to extend the analysis to situations where $m \geq 1$, this does not seem to be the empirically plausible parameter region, given the high rents in the financial sector. When $m \geq 1$ there is excess demand by informed dealers for good assets, so that dealers dissipate most of their informational rent through competition for good assets. Besides the fact that information may be too costly to acquire for most type 2 agents, there is a fundamental economic reason why $m < 1$ is to be expected in equilibrium. Indeed, even if enough type 2 agents have low costs $\varphi(d)$ so that if all of these agents became dealers we would have $m \geq 1$, this is unlikely to happen in equilibrium, as dealers would then compete away their informational rents to the point where they would not be able to recoup even their relatively low investment in dealer skills $\varphi(d)$.

We also restrict attention to parameter values for which dealers are not financially constrained in their purchase of a high quality asset in period 1. That is, we shall restrict ourselves to parameter values for which $p^d - 1 < \rho$. For this it is enough to assume that

$$\gamma\rho < 1 + \rho. \tag{7}$$

In addition, and in order to simplify the presentation in what follows, we restrict ourselves to situations where even in the absence of a dealer sector, $d = 0$, type 2 agents would prefer to become entrepreneurs and exercise the low effort rather than simply carry their endowments forward. We show in the appendix that to obtain this it is enough to assume that

$$\rho \left[1 + a_l \left(\gamma - 1\right)\right] > 1. \tag{8}$$

1.4 Definition of equilibrium

An equilibrium is given by: (i) prices p^* and p^{d*} in period 1 at which the organized and OTC markets clear; (ii) occupational choices by type 2 agents in period 0, which map into equilibrium measures of dealers d^* and entrepreneurs $(1 - d^*)$; (iii) incentive compatible effort choices a^* by entrepreneurs, which in turn map into an equilibrium matching probability $m(a^*, d^*)$; and (iv) type 2 agents prefer the equilibrium occupational choices rather than autarchy.

For simplicity, we restrict attention to symmetric equilibria in which all entrepreneurs choose the same effort in period 0. Given this assumption our economy admits two types of equilibria, which may co-exist. One is a *low-origination-effort equilibrium*, in which all entrepreneurs choose $a^* = a_l$. The other is a *high-origination-effort equilibrium*, in which all entrepreneurs choose $a^* = a_h$. This latter equilibrium is going to be the focus of what follows as it is only in this equilibrium that there is a social role for dealers. The main result of this paper is that whenever there is a role for informed dealers to support the high effort equilibrium there are "too many of them," in a sense to be made precise below. In what follows we sometimes refer to d^* as the size of the financial sector and thus when there are too many dealers we say that the financial sector is too big.

We begin by describing equilibrium borrowing and trading in assets in period 1, for any given occupation choices d^* of type 2 agents and any given action choices a^* of entrepreneurs in period 0. We are then able to characterize expected payoffs in period 0 for type 2 agents under each occupation. With this information we can then provide conditions for the existence of either equilibrium and present illustrative numerical examples.

2 Equilibrium payoffs and the moral hazard problem

In this section we derive the equilibrium payoffs associated with becoming and entrepreneur and a dealer, which determine the occupational choice. For this we first need to offer a minimal characterization of agents's actions along the equilibrium path at date 1, when trading occurs. In our framework we allow for collateralized lending at the interim date and thus the question arises as to whether agents in distress prefer to borrow rather than sell. We show in Lemma 1 that this is not the case. We also show that a patient entrepreneur that follows the equilibrium action prefers to keep his asset rather than sell it (Lemma 2). These two results are enough to yield the equilibrium expected payoffs, as of date 0, of either becoming and entrepreneur or a dealer. We then turn to the characterization of the entrepreneurs' moral hazard problem at date 0 and show conditions under which the high and low effort actions are incentive compatible.

2.1 Equilibrium borrowing and asset trading in period 1

We begin by describing behavior in period 1 in either the low or the high effort equilibrium. In period 1, d^*, a^* and, $m(a^*, d^*)$ are given. For any (a^*, d^*) :

Lemma 1 *In period* 1 *neither (a) an entrepreneur, nor (b) an impatient dealer ever borrows.*

Item (a) of this result follows immediately from our assumption that only safe collateralized borrowing is available to the entrepreneur. But this result holds more generally, even when risky borrowing is allowed. Indeed, in an asset sale the buyer obtains both the upside and the downside of the asset, while in a loan the lender is fully exposed to the downside, but only partially shares in the upside with the borrower. As a result the loan amount is always less than the price of the asset. And since the holder of the asset wants to maximize consumption in period 1 he is always better off selling the asset rather than borrowing against it.

While impatient entrepreneurs always prefer to sell their asset in period 1, the next lemma establishes that patient entrepreneurs never want to sell their asset.

Lemma 2 *Assume all entrepreneurs choose the same action. Then a patient entrepreneur (weakly) prefers not to put up his asset for sale in period* 1.

2.2 Equilibrium payoffs in period 0

We are now in a position to determine equilibrium payoffs for dealers and entrepreneurs in period 0. Since we examine incentive compatible symmetric equilibria, all entrepreneurs are treated identically; only dealers can differ since they may have different costs of acquiring information. Let $U(a|a', d)$ be the expected payoff of the entrepreneur who implements action a when all other entrepreneurs do a' and the measure of dealers is d. Similarly let $V\left(\widetilde{d}|a', d\right)$ be the expected payoff of dealer $\widetilde{d} \leq d$ when entrepreneurs implement action a' and the measure of dealers is d.

The entrepreneur's equilibrium expected payoff when the measure of dealers is $d < \bar{d}$ is

$$\begin{aligned} U(a^*|a^*, d) = -\psi(a^*) \quad &+ \quad \pi\left[a^* m(a^*, d)\, p^d(a^*, d) + (1 - a^* m(a^*, d))\, p(a^*, d)\right] \\ &+ \quad (1-\pi)\rho\left[1 + a^*(\gamma - 1)\right] \end{aligned} \tag{9}$$

where recall that

$$p^d(a^*, d) = \kappa\gamma\rho + (1-\kappa)\, p(a^*, d) \quad \text{with} \quad p(a^*, d) = \frac{\rho[a^*(1 - m(a^*, d))\gamma + (1 - a^*)]}{1 - a^* m(a^*, d)} \tag{10}$$

and $m(a^*, d)$ is given by

$$m(a^*, d) = \frac{d(1-\pi)}{a^*(1-d)\pi}. \tag{11}$$

In expression (9) the first term, $-\psi(a^*)$, is the cost of exercising effort a^*, which is 0 if $a^* = a_l$ and ψ if $a^* = a_h$. The first term in brackets is the utility of the entrepreneur if subject to a liquidity shock, which happens with probability π. If he draws a project yielding $\gamma\rho$, which occurs with probability a^*, and gets matched to a dealer, which happens with probability

$m(a^*, d)$, then he is able to sell the project for $p^d(a^*, d)$, the price for high quality projects in the dealers' market. If one of these two things does not occur, an event with probability $1 - a^* m(a^*, d)$, then the agent needs to sell his project in the uninformed exchange for a price $p(a^*, d)$. Finally, the second term in brackets is the utility of the entrepreneur conditional on not receiving a liquidity shock. The expressions for the prices and matching probabilities as a function of the measure of dealers for a given equilibrium level of effort a^* are given in expressions (10) and (11), respectively.

Let $V\left(\widetilde{d}|a^*, d\right)$ be the expected utility of the dealer $\widetilde{d} \leq d$ as a function of the measure of dealers d. Then

$$V\left(\widetilde{d}|a^*, d\right) = -\varphi\left(\widetilde{d}\right) + 1 + (1-\pi)(1-\kappa)(\rho\gamma - p(a^*, d)). \tag{12}$$

The first term in (12), $-\varphi(\widetilde{d})$, is agent $\widetilde{d}$'s cost of acquiring information, the second is the agent's endowment and the third is the surplus that the dealer obtains in the absence of a liquidity shock, which happens with probability $1-\pi$, as in this case the agent captures a fraction $1-\kappa$ of the difference between the good asset's payoff, $\gamma\rho$ and the price at which assets trade in the exchange, $p(a^*, d)$.

The next proposition provides a characterization of both $U(a^*, d)$ and $V\left(a^*, d \mid \widetilde{d}\right)$ as a function of the measure of dealers d.

Proposition 3 *(a) The utility of an entrepreneur is a decreasing and concave function of the measure of dealers,* d*, and (b) the utility of dealer* $\widetilde{d}$ *is an increasing and convex function of the measure of dealers,* d*.*

To better understand the previous proposition it is useful to consider first the following result, which is immediate,

Proposition 4 *(a) The matching probability* $m(a, d)$ *is an increasing and convex function of the measure of dealers and (b) the price in the uninformed exchange* $p(a, d)$ *is a decreasing and concave function of the measure of dealers; moreover* $p(a_l, d) < p(a_h, d)$*.*

(a) is obvious, but (b) is at the heart of our results. As the number of dealers increases entrepreneurs with good projects are more likely to get matched with some dealer. This can only come at the expense of worsening the pool of assets flowing into the uninformed exchange, which leads to lower prices there. In other words, dealers in the OTC market *cream skim* the good assets and thereby impose a *negative externality* on the organized market. Cream skimming thus improves terms for dealers in the OTC market and worsens them for entrepreneurs in distress.

The intuition behind Proposition 3 follows from the previous logic. Start with the dealers' expected payoffs. The larger their measure, the lower the price of the asset in the uninformed exchange and thus the higher the surplus that accrues to them, $(1-\kappa)(\gamma\rho - p(a^*, d))$ when they acquire high quality assets from entrepreneurs in distress at date 1. This results in an increasing expected payoff for the dealers as a function of d, holding fixed the action of entrepreneurs. The additional rents that accrue to dealers when their measure increases can only come at the expense of the entrepreneurial rents. It follows that the entrepreneur's expected payoff is a decreasing function of d.

That the entrepreneur's expected payoff is a decreasing function of d is a more subtle result than may appear at first. Indeed notice that an increase in the number of dealers has two effects on the utility of the entrepreneurs. On the one hand, if a good project is drawn, the probability of being matched with an informed dealer goes up, which benefits the entrepreneur as he obtains a better price from the dealer than from the exchange. But an increase in the number of dealers results in more *cream skimming* and thus in lower prices in the uninformed exchange, which in turn leads dealers to bid less for the asset in OTC markets. Overall, all entrepreneurs in distress are hurt, whether they get matched or not with an informed dealer. Proposition 3 establishes that the latter effect overwhelms the first positive effect yielding a decreasing utility for the entrepreneur as a function of the measure of dealers in the economy. This result captures somewhat the populist sentiment of *Main street* towards *Wall street*, as a large financial sector can only come at the expense of the profits of entrepreneurs.

Another implication of our model is that dealers also prefer dealing in market equilibria with low quality origination of assets. The reason is that for a given number of dealers, the price in the exchange is lower the lower the proportion of good projects generated, as the same amount of cream skimming results in fewer good projects flowing into the exchange. Thus if dealers could induce more bad asset origination, they would do so.

In the next section we study the moral hazard problem of entrepreneurs and show that a strictly positive measure of dealers is needed to support an equilibrium with high effort.

2.3 Entrepreneur moral hazard

A necessary condition for any symmetric equilibrium is that it is *incentive compatible* for entrepreneurs to choose the equilibrium effort, that is,

$$U(a^*|a^*, d^*) \geq U(a|a^*, d^*) \qquad \text{for} \qquad a \neq a^*. \tag{13}$$

Throughout we write $U_h\left(d\right)$ for the equilibrium expected payoff of the entrepreneur along the high effort equilibrium path as a function of d and denote by $U_{hl}\left(d\right)$ the utility of the entrepreneur that deviates and implements action a_l instead of a_h, that is,

$$U_h\left(d\right) = U(a_h|a_h, d) \qquad \text{and} \qquad U_{hl}\left(d\right) = U(a_l|a_h, d),$$

where the subscript hl refers to the payoff from a *deviation* from a_h to a_l. A similar notation simplification applies when $a^* = a_l$.

Consider first incentive compatibility in the high effort equilibrium, where all entrepreneurs choose a_h. Recall that the entrepreneur's expected payoff in period 0 when choosing effort a_h in the high effort equilibrium as a function of the measure of dealers is given by:

$$U_h\left(d\right) = -\psi + \pi\left[a_h m_h\left(d\right) p_h^d\left(d\right) + \left(1 - a_h m_h\left(d\right)\right) p_h\left(d\right)\right] + (1-\pi)\rho\left[1 + a_h\left(\gamma - 1\right)\right], \tag{14}$$

where $p_h^d\left(d\right)$, $p_h\left(d\right)$ and $m_h\left(d\right)$ refer to the prices and matching probabilities.

Suppose now that an entrepreneur chooses to *deviate* in period 0 by choosing the low effort a_l. In this case, as Proposition A in the appendix states, it is optimal for this entrepreneur to put his asset for sale in the OTC market even when he is not hit by a liquidity shock. Indeed assume that this is the case. If the entrepreneur receives a bid from one of the informed dealers he rationally infers he has a good asset, refuses the bid and instead carries it to maturity. If instead he does not receive a bid it may be because he drew a good project but did not get matched to a dealer or because the project is indeed bad and thus dealers do not bid for it. In either case the agent lowers his posterior on the quality of his asset. This private valuation is always below the average quality of projects flowing to the uninformed exchange. The reason is that this pool is relatively good, as the rest of the entrepreneurs implemented the high effort. Thus, the shirking entrepreneur if not found by a dealer, sells at the exchange, hiding behind the better projects of entrepreneurs that chose high effort. More formally, Proposition A shows that the payoff of an entrepreneur that deviates to the low effort when the measure of dealers is given by d is,

$$U_{hl}(d) = p_h(d) + a_l m_h(d)(\gamma\rho - p_h(d))(\pi\kappa + (1-\pi)). \tag{15}$$

High effort is incentive compatible if, and only if, $U_h(d) \geq U_{hl}(d)$. Denote by $\Delta U_h(d)$ the difference in expected *monetary payoffs*, not accounting for the effort cost ψ, from the high versus the low effort when the measure of dealers is d:

$$\begin{aligned} \Delta U_h(d) &= \psi + U_h(d) - U_{hl}(d) \qquad (16)\\ &= \pi\Delta a m_h(d)\kappa(\gamma\rho - p_h(d)) \\ &+ (1-\pi)\left[\rho(1 + a_h(\gamma - 1)) - (p_h(d) + a_l m_h(d)(\gamma\rho - p_h(d)))\right]. \end{aligned}$$

Incentive compatibility requires that

$$\Delta U_h(d) \geq \psi. \tag{17}$$

Now consider incentive compatibility in the low effort equilibrium, where all entrepreneurs choose a_l. In this case, an entrepreneur's expected payoff in period 0 along the equilibrium path is:

$$U_l(d) = \pi\left[a_l m_l(d) p_l^d(d) + (1 - a_h m_l) p_l(d)\right] + (1-\pi)\rho\left[1 + a_l(\gamma - 1)\right] \tag{18}$$

where $p_l^d(d)$, $p_l(d)$, and $m_l(d)$ are defined as in the previous case with the obvious changes in notation.

We show in Proposition A in the appendix that an entrepreneur who chooses to *deviate* from this equilibrium in period 0 by exercising the high effort a_h is better off holding on to his asset until period 2, unless he is hit by a liquidity shock. The reason is that now his private valuation is higher than the average quality of the assets in the exchange. Proposition A states that his expected payoff under the deviation is given by:

$$U_{lh}(d) = -\psi + \pi \left[p_l(d) + a_h m_l(d) \kappa (\gamma\rho - p_l) \right] + (1-\pi)\rho \left[1 + a_h (\gamma - 1) \right].$$

Incentive compatibility in the low effort equilibrium when the measure of dealers is d again requires that $U_l(d) \geq U_{lh}(d)$, or if we define again $\Delta U_l(d)$ as the difference in expected *monetary payoffs* (not accounting for effort costs ψ) between the utility under the deviation and the utility that obtains if the agents sticks to the candidate equilibrium action a_l:

$$\begin{aligned} \Delta U_l(d) &= \psi + U_{lh}(d) - U_l(d) \\ &= \pi \Delta a m_l(d) \kappa (\gamma\rho - p_l(d)) + (1-\pi)\rho\Delta a (\gamma - 1), \end{aligned}$$

then incentive compatibility requires that

$$\Delta U_l(d) \leq \psi. \tag{19}$$

The next proposition characterizes the functions $\Delta U_h(d)$ and $\Delta U_l(d)$.

Proposition 5 *(a) $\Delta U_h(d)$ and $\Delta U_l(d)$ are both strictly increasing functions of d and (b) $\Delta U_h(d) < \Delta U_l(d)$ for all $d \geq 0$.*

The functions $\Delta U_h(d)$ and $\Delta U_l(d)$ are shown in Figure 2. The reason why these functions are increasing in the mass of dealers d is simply that with a greater mass of dealers there is a greater likelihood $m(a^*, d)$ for an entrepreneur with a good asset to be matched with an informed dealer. Thus, an entrepreneur deviating from a low-origination equilibrium a_l by choosing a_h is more likely to get rewarded with a match in the OTC market in the event that he has a good asset. Therefore his incremental payoff from deviating is larger. For an entrepreneur deviating from a high-origination equilibrium a_h by choosing a_l, the higher is d the more good assets get skimmed in the OTC market, which results in a lower price p in the organized market at which the entrepreneur can sell his bad asset. This is why $\Delta U_h(d)$ is also increasing in d.

Item (b) is a result of the different out-of-equilibrium behavior of entrepreneurs that deviate from the high effort equilibrium and entrepreneurs that deviate from the low effort equilibrium. When all entrepreneurs choose high effort the deviant agent has "more options" than when all entrepreneurs choose low effort. A deviant entrepreneur who implements a_l instead of a_h can benefit from selling in the uninformed exchange, even in the absence of a liquidity shock, because his private valuation is lower than the average quality of the assets being traded. This is not the case in the low effort equilibrium; a deviant entrepreneur implements a_h and if he sells his asset in the uninformed exchange in the absence of a liquidity shock (and a match in the OTC market) he would be providing a subsidy rather than receiving it. It follows that the deviation when entrepreneurs implement a_h is more profitable than when they implement a_l, and thus $\Delta U_h(d) < \Delta U_l(d)$.

Next, if we define $\widehat{d}_h$ and $\widehat{d}_l$ by

$$\widehat{d}_h = \inf\{d \leq \bar{d} : \Delta U_h(d) \geq \psi\} \qquad \text{and} \qquad \widehat{d}_l = \sup\{d \leq \bar{d} : \Delta U_l(d) \leq \psi\} \tag{20}$$

we are able to establish:

Proposition 6 *(a)* $\widehat{d}_l \leq \widehat{d}_h$. *(b) A low effort equilibrium can only be supported for* $d \in [0, \widehat{d}_l]$. *(c) A high effort equilibrium can only be supported for* $d \in [\widehat{d}_h, \bar{d}]$ *and* $\widehat{d}_h > 0$.

Proposition 6 is key in establishing the main results of the paper and merits emphasizing some of its implications. In Figure 2 we consider two possible costs of exercising the high effort, ψ and ψ'. If the high effort is socially optimal, and we provide a condition below under which this is the case, then the existence of an OTC market of at least size $\widehat{d}_h$ is necessary to support it. Even when the cost of exercising the high effort is arbitrarily small this effort level is never incentive compatible when d is close to 0. The reason is that, under the candidate high effort equilibrium, the price of the asset in the uninformed exchange is very high when d is close to 0. There is a large measure of entrepreneurs, $1 - d$, all exercising the high effort and there is little cream skimming and hence the quality of the pool of assets flowing into the exchange is high. Thus the price in the uninformed exchange is close to $[1 + a_h (\gamma - 1)] \rho$, the price the asset commands in the absence of any cream skimming. An agent deviating to low effort, if not receiving an offer from an informed dealer, will be able to sell the asset at $t = 1$, independently of whether he suffers a liquidity shock, for a price higher than his uninformed private valuation.[14] Also because there are few informed dealers the entrepreneurs have little hopes of being matched to them at date 1 and thus of capturing some of the surplus $\gamma\rho - p(d)$; thus, given that his high effort provision is likely to go unrewarded in case of distress, the agent prefers simply to save on effort costs and free ride on the large pool of entrepreneurs exercising the high effort.

A second implication of Proposition 6 is that a low effort equilibrium fails to exist for a sufficiently low cost of providing the high effort, as happens when this cost is ψ' in Figure 2. When entrepreneurs are choosing low effort, the price in the uninformed exchange is low. Thus, if effort is not very costly, an entrepreneur prefers to exercise the high effort and get rewarded in the state in which he draws the high quality project and suffers no liquidity shock. In addition when $d > 0$ he will be matched to an informed dealer in case of a liquidity shock if he has a good project. These two effects are increasing in Δa. Indeed, as is apparent in Figure 2, the range of ψs for which a low effort equilibrium does not exist is increasing in Δa.

[14] And keep the asset if he obtains a bid from an informed dealer and is not subject to a liquidity shock, for in this case he learns the asset will yield $\gamma\rho$ at date 2.

3 Allocation of talent and welfare

3.1 The equilibrium size of the financial and real sectors

We now turn to a central question of our analysis: What is the optimal allocation of talent to the financial sector? Is there too much information acquisition in financial markets? In our model, these questions boil down to determining whether the equilibrium measure of dealers d^* is too large. As we saw in Proposition 6, a low effort equilibrium can only be supported when $d \leq \widehat{d}_l$, and high effort equilibrium can only be supported if $d \geq \widehat{d}_h$. Low effort equilibria thus are associated with relatively small financial sectors when compared with high effort equilibria.

It is relatively simple to construct examples for which there is no symmetric equilibrium and for which there are multiple ones. Rather than provide a full characterization of the many possible cases, we provide in what follows examples of three possible cases: One in which there are only high effort equilibria, one in which there are only low effort equilibria and one in which low and high effort equilibria coexist. Recall also that for a particular (a^*, d^*) to be an equilibrium a^* must be incentive compatible and, given (21), d^* has to be such that

$$\begin{aligned} U\left(a^*|a^*, d^*\right) &\geq V\left(d|a^*, d^*\right) \qquad \text{for} \qquad d \geq d^* \\ U\left(a^*|a^*, d^*\right) &< V\left(d|a^*, d^*\right) \qquad \text{for} \qquad d < d^*. \end{aligned}$$

In the examples, the cost of acquiring information is simplified to a step function:

$$\varphi\left(d\right) = \overline{\varphi} \quad \text{for} \quad d < \overline{d} \qquad \text{and} \qquad \varphi\left(d\right) = +\infty \quad \text{for} \quad d \geq \overline{d}, \tag{21}$$

Under (21) all dealers have identical costs and thus when plotting the expected payoff function of one of them we also plot that of the *marginal* dealer, who determines the size of the OTC market. We may thus define $V\left(a, d\right) := V(\widetilde{d}|a, d)$, for any $\widetilde{d} \leq d < \bar{d}$.

3.1.1 High effort equilibria

Consider the following parameter values

$$a_h = .75 \quad a_l = .55 \quad \gamma = 1.5 \quad \rho = .8 \quad \kappa = .25 \quad \pi = .5. \tag{22}$$

We also choose

$$\psi = .001 \qquad \overline{\varphi} = 0 \qquad \text{and} \qquad \overline{d} = .35.$$

In this case, inequality (6) holds since $m < 1$ for $d = \bar{d}$. There is no low effort allocation that is incentive compatible in this example since $(1-\pi)\rho\Delta a\,(\gamma - 1) > \psi$. High effort is incentive compatible if $d \geq \widehat{d}_h = .0536$. There are two high effort equilibria and they are shown in Figure 3. There is an unstable equilibrium with $d_1^* = .3106$ in which all agents $d \leq \overline{d}$ are indifferent between becoming entrepreneurs or dealers. There is also a stable equilibrium with $d_2^* = \overline{d} = .35$, in which dealers are *strictly* better off than entrepreneurs. Notice that all agents who can become dealers at a finite cost are dealers in this equilibrium.

The price of assets in the OTC market in the unstable equilibrium is $p^d\,(a_h, d_1^*) = 1.0180$, so that a dealer needs some leverage in order to finance the purchase of the asset. In the stable equilibrium leverage is not needed as $p^d\,(a_h, d_2^*) = .9833 < 1$.

3.1.2 Low effort equilibria

Suppose (22) holds but

$$\psi = .0475 \qquad \overline{\varphi} = .06 \qquad \text{and} \qquad \overline{d} = .15.$$

Here $\widehat{d}_h = \infty$, and there are no high effort equilibria, and $\widehat{d}_l = \bar{d}$. As shown in Figure 4, there are three (low effort) equilibria. There is a stable equilibrium where $d_1^* = 0$. Indeed, when there are no dealers $U(a_l \mid a_l, 0) > V(0 \mid a_l, 0)$. There is also an unstable equilibrium with $d_2^* = .0781$ and $U(a_l \mid a_l, .0781) = V(.0781 \mid a_l, .0781)$, that is, the marginal dealer is indifferent between being a dealer or an entrepreneur. Finally, there is a stable equilibrium with $d_3^* = .15$ where $U(a_l \mid a_l, .15) < V(.15 \mid a_l, .15)$.

3.1.3 Coexistence of high and low effort equilibria

Suppose now that

$$\kappa = .5, \qquad \psi = .0410 \qquad \overline{\varphi} = .03 \qquad \text{and} \qquad \bar{d} = .41,$$

and the rest of the parameters are as in (22). There are three equilibria, two that feature low effort and one stable high effort equilibrium.

Then $\widehat{d_l} = .0545$ and there are two low effort equilibria. A stable one with $d_1^* = 0$, since $U(a_l|a_l, 0) > V(0, a_l, 0)$, and an unstable equilibrium where type 2 agents with $d \leq \bar{d}$ are indifferent between becoming dealers or entrepreneurs and $d_2^* = .05$.

In this example $\widehat{d_h} = .4020$. The allocation $(a_h, d_3^* = .41)$ is a stable high effort equilibrium.

3.2 Welfare: Are OTC markets too large?

3.2.1 Constrained efficiency: Definition

Our notion of *constrained efficiency* is based on the standard idea that the social planner should not have an informational advantage relative to an uninformed market participant. Thus, we allow the planner to dictate the occupation of type 2 agents but we do not let the planner make any decisions based on the information obtained by informed dealers. Given a vector of parameters $\mathcal{A} = (a_h, a_l, \gamma, \rho, \kappa, \pi, \psi, \bar{d})$ and the cost function $\varphi(d)$ the planner chooses d knowing that trade will occur in time 1 in the OTC market with d dealers and in the organized exchange at equilibrium prices. The planner's problem in period 0 is then to pick the measure d of type 2 agents that maximizes ex-ante social surplus. Since type 1 agents get no surplus in equilibrium the planner only has to weight the utility of type 2 agents, and we assume that all type 2 agents receive equal weight. If the planner wishes to implement low effort, the optimal choice is obviously $d = 0$ which yields a total surplus that equals $\rho(1 + a_l(\gamma - 1))$. If the planner chooses to implement high effort, she must choose a $d \geq \widehat{d}_h$ and this yields surplus:

$$[\rho (1 + a_h (\gamma - 1)) - \psi] (1 - d) - \int_0^d \varphi (u) \, du, \tag{23}$$

which is monotonically decreasing in d and thus the optimal choice is $d = \widehat{d}$.

3.2.2 The allocation of talent and constrained efficiency

We focus on situations where there is a role for the financial sector. The high effort is socially efficient if

$$[\rho (1 + a_h (\gamma - 1)) - \psi] \left(1 - \widehat{d}_h\right) - \int_0^{\widehat{d}_h} \varphi (u) \, du \geq \rho (1 + a_l (\gamma - 1)) \, . \tag{24}$$

The first term of (24) is the output produced by the $1 - \widehat{d}_h$ entrepreneurs when they implement the high effort, net of costs. The integral corresponds to the information acquisition costs of type 2 agents who become dealers. The high effort is socially efficient if this term is more than what society would obtain if all type 2 agents become entrepreneurs and perform the low effort, which by (8) dominates the allocation where type 2 agents prefer to simply carry their endowment to subsequent dates. Of course, if a high effort equilibrium exists, it is unlikely that $d_h^* = \widehat{d}_h$. The next proposition states this fact more precisely.

Proposition 7 *Suppose that it is socially efficient to implement the high effort action; that is, inequality (24) holds. Then given any vector of parameters $\mathcal{A}$ and any $\epsilon > 0$, there exists a vector $\mathcal{A}'$ with $|\mathcal{A}' - \mathcal{A}| < \epsilon$ such that for the parameter vector $\mathcal{A}'$ all equilibria are inefficient and any high effort equilibrium features too many dealers in OTC markets.*

This Proposition does not rule out the possibility that an equilibrium involving low effort obtains when it is optimal to implement the high effort. In this case, in the (inefficient) low effort equilibrium there are too few dealers. In this equilibrium dealers receive too little compensation and only those with very low cost of becoming dealers, if any, choose to do so.

It is straightforward to verify that for the parameter values given in section 3.1.1 above the socially efficient origination effort is a_h:

$$[\rho(1 + a_h(\gamma - 1)) - \psi]\left(1 - \widehat{d}_h\right) - \varphi\widehat{d}_h - \rho(1 + a_l(\gamma - 1)) = .0201,$$

and thus both equilibria are inefficient and feature an excessively large financial sector in the form of a large measure of informed dealers. The intuition is by now clear. Conditional on a_h being efficient, the planner wants to support this level of effort with the minimum measure of dealers $\widehat{d}_h$, for adding "one" additional dealer detracts from productive entrepreneurial activities and does not improve incentives; but this level can only be supported as an equilibrium for a set of economies of measure zero. The reason is by now well understood: Entry into OTC markets creates a positive externality among dealers via the cream skimming and this leads to a larger OTC market than constrained efficiency would have it.

In the example in section 3.1.2, the constrained efficient allocation calls for a_l and $d = 0$. Notice that in that case there were three equilibria, two of which feature excessively large OTC markets and one that indeed supports the constrained social optimum, $(a^* = a_l, d_1^* = 0)$.

In the example in section 3.1.3 (24) is not met and thus high effort is not socially efficient, though it can be supported as a stable equilibrium. There is also an efficient low effort equilibrium with no financial sector and an inefficient one with a strictly positive measure of dealers.

The argument above highlights that, conditional on a particular level of effort, the different equilibria can be Pareto ranked in decreasing order of the measure of dealers. Thus in the example in section 3.1.1, the most efficient equilibrium is the unstable one, d_1^*, which dominates the stable one d_2^*. In the example in section 3.1.2, which deals with the low effort equilibria case, the result is that $d_1^* \succ d_2^* \succ d_3^*$. We summarize this discussion in the following proposition.

Proposition 8 *Equilibria with the same effort can be ranked by total ex-ante social surplus in decreasing order of the measure of dealers that OTC markets attract.*

4 Competition between dealers

In our model we assumed that an entrepreneur's bargaining power κ is invariant to the number of dealers, d. A plausible alternative assumption is that as the number of dealers increases so does the entrepreneurs' bargaining power. That is, $\kappa(d)$ is an increasing function of d: $\kappa' > 0$. In this section we show that the main results of the paper still hold under this generalization. In particular, Proposition 7, our main result, remains unaffected: If there is a social role for dealers in supporting the high effort *all* equilibria are generically inefficient and moreover any high effort equilibrium features inefficiently large OTC markets.

First notice that Proposition 5 remains valid when $\kappa' > 0$. In fact, in this case, the derivative of $\Delta U_h(d)$ with respect to d gains a single extra term

$$\kappa'(d)\pi\Delta a m_h(d)(\gamma\rho - p_h(d)) > 0.$$

Similarly, the derivative of $\Delta U_l(d)$ with respect to d gains a single positive extra term, with m_ℓ and p_ℓ replacing m_h and p_h respectively. Hence item (a) in Proposition 5 holds and, since $\Delta U_h(d) < \Delta U_l(d)$ for any κ, (b) follows as well.

Proposition 6, which describes the set of possible measures of dealers in the low and high effort equilibria, is a Corollary to Proposition 5 and thus holds as well when $\kappa' > 0$. This Proposition lies at the heart of the analysis in Section 4. Proposition 3 on the other hand no longer holds; the positive externality may be offset by the effect of greater competition on κ. But the monotonicity of each dealer's utility with respect to the measure of dealers is unrelated to our main result. For instance, if a high effort equilibrium exists, it has to generically feature a measure of dealers that is *strictly* greater than $\widehat{d}_h$, which is *inefficient*. Proposition 7 thus still holds.[15]

[15]Intuition suggests that our main results also hold for the implausible case where $\kappa' < 0$. If an increase in the number of dealers increases the dealers bargaining power, dealers benefit from double cream skimming. The reservation prices *and* the bargaining power of entrepreneurs go down as dealers enter.

5 Discussion and applications

5.1 Cream skimming in financial markets

In addition to the example of futures and forwards discussed in the introduction, there are several other examples of cream-skimming in financial markets. Perhaps the most direct illustration concerns the rise of private securities markets. One of the main goals of the Securities Act of 1933 was to protect unsuspecting investors against fraud, via registration of securities offered to the public and other reporting standards. The Securities Act, however, allowed for the possibility of non-registration "for transactions by an issuer not involving any public offering." Successive rulings and clarifications led to Rule 506 according to which an offering is exempt from securities regulation as long as takers are limited to accredited investors and no more than 35 nonaccredited investors.[16] This exemption facilitated the rise of a private placement market, which went from \$5bn raised in 1980 to \$250bn in 2006. This market, in turn, funded the rise of the VC and private buyout industry. Restrictions on the resale of securities, however, prevented securities underwriters from taking advantage of the exemption, as the Securities Act prohibited the resale of any unregistered securities unless the sellers were persons other than the issuer, underwriter, or dealer. But, in 1972 Rule 144 somewhat weakened these restrictions and introduced substantial flexibility by imposing a holding period requirement for resale (of six months) instead of an outright ban on resale.

A watershed moment in the evolution of the private securities market came in 1990, when the SEC adopted Rule 144A which provides a safe harbor from the registration requirements of the Securities Act of 1933 for certain private resales of minimum \$500,000 of restricted securities to QIBs (qualified institutional investors), generally large institutional investors with \$100m in investable assets. The adoption of Rule 144A greatly increased the

[16]SEC Rule 501(a) defines an accredited investor as any natural person whose individual net worth, or joint net worth if said natural person has a spouse, exceeds \$1.000.000 at the time of the purchase; there are similar requirements on income. A nonaccredited investor instead has to be "sophisticated" in that he has to have sufficient knowledge and experience to evaluate the merits of the offering independently.

liquidity of private securities as it facilitated trading amongst QIBs. Effectively the 144A market allows sophisticated investors to freely trade private securities without any registration requirements. Rule 144A, as well as the Sarbanes-Oxley Act, led to a remarkable increase in private equity issuance. In 2006 more equity capital was raised via Rule 144A private placements ($162bn) than in IPOs in Amex, NASDAQ and NYSE ($154bn).[17]

Financial intermediaries rushed to design proprietary platforms where QIBs trade Rule 144A shares. Goldman Sachs created the Goldman Sachs Tradable Unregistered Equity (GSTrUE) platform, Citi, Lehman, Merrill Lynch, BoNY and Morgan Stanley created Opus-5 and NASDAQ followed suit with Portal. An important milestone in the development of this market occurred when Oaktree Capital Management LLC sold an equity stake for $800m.[18]

There are many reasons for the success of private equity placements. Escaping the regulatory burdens associated with Sarbanes-Oxley must be an important factor, but in addition there is some evidence that higher quality issuers are flocking to Rule 144A rather than public offerings. This trend has led some to argue that exchanges run the risk of being deprived of high quality issues. For instance, Roger Ehrenberg, former CEO of Deutsche Bank's hedge fund platform DB Advisors, said:

> I think they (Rule 144A equity issuances) will quickly detract from the Nasdaq, NYSE and Amex. These private exchanges will effectively **skim the cream off the market**. The very highest quality issuers will forgo the public markets to issue on the private exchanges. (quoted in Lambe (2007, page 42))

The liquidity and transparency provided by these platforms are an important reason for the success of Rule 144A issues, but the fact that participants are restricted to be QIBs leads to cream skimming: High quality issuers have an incentive to raise equity capital in Rule 144A

[17] See Lambe (2007, page 40) and Tang (2007, Figure 1) for a figure showing private equity capital issuance compared to IPO issuance for the period 2002 to 2006.

[18] See Sjostrom (2008) for a description of this deal as well as the discussion of Rule 144A in general. For a legal analysis that is contemporaneous with the 1990 adoption of Rule 144A see Testy (1990).

markets rather that through IPOs because QIBs are better informed than the average IPO buyer and can thus offer some price improvement. But one consequence is a lower quality of offerings in public markets.

5.2 Some evidence on the allocation of talent

The fundamental inefficiency in our model arises because too many agents become dealers instead of entrepreneurs. Anecdotal evidence suggests that in the last two decades the brightest minds found a way to Wall Street in either the asset management industry or the broker-dealers and investment banks. Golding and Katz (2008) find evidence of this shift using the Harvard and Beyond data set:

> The most striking changes with regard to occupations concern the ascendancy of finance and management. Amongst the oldest cohort (C1970), 22 percent of the men were in occupations in these fields 15 years after their class graduated. But for the youngest cohort (C1990), 38 percent were. The change moreover was driven primarily by positions in finance, which increased from 5 to 15 percent of the total. The relative growth in business occupations for men came largely at the expense of those in law and medicine, which declined from 39 to 30 percent of the total [...] 12 percent of women in C1970 were in management and finance occupations, but 23 percent were in C1990.[19]

5.3 On the inevitability of dealer based markets

Our model offers a simple theory where OTC markets arise naturally, even in the presence of well functioning exchanges. Both sides of the market have an incentive to meet outside the exchange: Entrepreneurs in distress with good projects may get recognized as such by informed dealers and thus obtain better prices for their assets than they would in the exchange

[19]Goldin and Katz (2008, page 366.)

and dealers can use their information to cream skim good projects. Our paper offers a novel theory of endogenous financial markets: Informed agents will have an incentive to create dealer based markets in which to lever their superior information and cream skimming creates a *positive externality* that induces too much information acquisition. This is in contrast with Grossman and Stiglitz (1980) where, instead, too little information is produced as some of it is imputed in prices and thus those bearing the costs of acquiring information do not fully capture the returns associated with it. Our point is that in Grossman and Stiglitz (1980) informed agents can only trade in organized exchanges when in reality they often create parallel dealer based markets where they lever their information without having this information revealed in prices.

In our model all firms that have a good project have exactly the same probability of being found by a dealer. If the probability of obtaining a good project or being matched to a dealer differs across entrepreneurs, those with more favorable prospects would gain more from the presence of dealers. As the probability of being matched when having a good project converges to 1, entrepreneurs would adopt high effort, independently of the effort done by others and when faced with a liquidity shock benefit from trading directly with a dealer. Thus it is not surprising that some "big name" firms, which have long term relationships with banks, also lobby for keeping OTC markets in their present form.[20]

In our model dealers profit from the opaqueness of OTC transactions. This explains efforts of dealers to prevent OTC contracts from being transferred to organized platforms,[21]

[20]See Scannell (2009), who writes "Companies from Caterpillar Inc. and Boeing Co. to 3M Co. are pushing back on proposals to regulate the over-the-counter derivatives market, where companies can make *private deals* to hedge against sudden moves in commodity prices or interest rates." (Emphasis ours).

[21]The furious lobbying activity of some banks, as well as the ISDA on their behalf, to avoid any major changes in the organization of OTC markets has been amply documented in the press. See for example Leising (2009), Morgenson (2010) and Tett (2010). In fact centralized clearing seems to be *less* of a problem for dealers than execution. For instance, Harper, Leising, and Harrington (2009) write:" [T]he banks ... are expected to lobby to remove any requirements that the contracts be executed on exchanges because that would cut them out of making a profit on the trades, according to lawyers working for the banks."

and why the largest Wall Street firms seem so intent on avoiding disclosure of prices and fees.[22]

5.4 IT and the growth of compensation in the financial industry

Figure 1 shows that the growth in median compensation in the financial industry is driven by the broker-dealers, which constitute the main entry in 'other finance'. Broker-dealers are the main players in OTC markets, and the units inside commercial banks and insurance companies, such as AIG's infamous Financial Products group, that have been richly rewarded during the boom years were present in these markets. But the timing of the abnormal growth in 'other finance' requires an explanation.

Philippon and Resheff (2008), amongst others, argue that it is the wave of deregulation that led to the phenomenal profits and growth of the financial services industry. This, undoubtedly played a role but our model suggests a different hypothesis: that improvements in information technology (IT) have decreased the costs of processing and organizing financial information in a way that has particularly benefitted OTC markets, where information traditionally was dispersed and hard to obtain. In our model this IT revolution could be captured by an increase in the maximum measure of dealers, $\overline{d}$. Consider the example in section 3.1.1, which featured a single stable high effort equilibrium. An improvement in IT can be captured by the number of type 2 agents for whom $\varphi(d) = 0$, which goes from $\overline{d}$ to $\overline{d} + \varepsilon$. This would lead to a new stable high effort equilibrium with a larger OTC market (by an amount ϵ) higher profits for dealers present in the market (both entrants and incumbents) and lower ex-ante profits for entrepreneurs.

[22] See for example Story (2010), who reports on the efforts by the largest banks to thwart an initiative by Citadel, the Chicago hedge fund, to set up an electronic trading system that would display prices for CDSs.

6 Conclusions

We have presented a model of occupational choice where agents can choose between becoming entrepreneurs and engaging in productive activities or acquiring information and becoming dealers. We identify a novel externality, cream skimming in OTC-like markets, that leads to inefficiencies in financial markets. In particular we show that this externality leads to excessive profits in the financial sector. Moreover, if one believes that there is a social role for financial markets in mitigating moral hazard problems at origination, then we show that the financial markets that arise in equilibrium are *always* too large.

Our theory helps explain the rise in compensation in the financial services industry and why is it concentrated among some financial entities and not others. We argue that it is the intermediaries which are present in OTC markets, mainly broker-dealers and the broker-dealer arms of large commercial banks, that capture these excessively large rents, which is consistent with observed trends in financial markets. In addition, our framework rationalizes the lobbying efforts of, not only banks, but also corporations on the other side of the market to preserve OTC markets in their current form, as was observed during the run-up to the passage of the Dodd-Frank Act in July of 2010 on financial regulatory reform. To the extent that opaque OTC markets facilitate informational rent extraction by informed dealers, it is to be expected that the move of derivatives and swap trading onto organized exchanges and clearinghouses required by the Dodd-Frank Act of 2010 will be resisted by the financial industry. We expect that a first line of defense is likely be to over-customize derivatives contracts and to offer fewer standardized, plain-vanilla contracts (which will be required to trade on organized exchanges). The second line of defense is likely to be to set up clearinghouses that maintain opacity and do not require disclosure of quotes. A third line of defense is to ensure that the operation of clearinghouses remains under the control of the main dealers, as has been suggested by Story (2010).[23]

[23] See Story (2010).

REFERENCES

Afonso, Gara (2010) "Liquidity and Congestion," *Journal of Financial Intermediation,* forthcoming.

Allen, Franklin and Douglas Gale (2000), *Comparing Financial Systems*, The MIT Press, Cambridge, Ma.

Baumol, William J. (1990), "Entrepreneurship: Productive, Unproductive, and Destructive,"*Journal of Political Economy*, Vol. XCIII, 893-921.

Bernanke, Ben and Mark Gertler (1989) "Agency Costs, Net Worth and Business Fluctuations,"*American Economic Review, Vol. 79, pp. 14-31.*

Binmore, Ken, Avner Shaked and John Sutton (1989) "An Outside Option Experiment,"*Quarterly Journal of Economics*, Vol. 104, No. 4, 753-770.

Binmore, Ken, Ariel Rubinstein and Asher Wolinsky (1986) "The Nash bargaining Solution in Economic Modeling,"*Rand Journal of Economics*, XVII, 176-88.

Duffie, Darrell, Nicolae Garleanu and Lasse H. Pedersen (2005) "Over-the-Counter Markets," *Econometrica,* vol. 73 (6), pp 1815-1847.

Glode, Vincent, Richard Green and Richard Lowery (2010) "Financial Expertise as an Arms' Race,"Wharton School Working Paper.

Goldin, Claudia and Larry Katz (2008) "Transitions: Career and Family Life Cycles of the Educational Elite,"*American Economic Review Papers & Proceedings* May, 98:2, 363-369.

Grossman, Sanford and Joseph Stiglitz (1980) "On the Impossibility of Informationally Efficient Markets,"*American Economic Review* Vol. 70, pp. 393-408.

Harper, Christine, Matthew Leising and Shannon Harrington (2009) "Wall Street Stealth Lobby Defends $35 billion Derivatives Haul," Bloomberg, August 30th, 2009.

Holmstrom, Bengt and Jean Tirole (1997) "Financial Intermediation, Loanable Funds and the Real Sector,"*Quarterly Journal of Economics*, Vol. 112, 663-91.

Lagos, Ricardo and Gullaume Rocheteau (2009) "Liquidity in Asset Markets with Search Frictions," *Econometrica,* Vol. 77, No. 2, March, pp. 403-426.

Lagos, Ricardo, Guillaume Rocheteau and Pierre-Olivier Weill (2010) "Crises and Liquidity in Over the Counter Markets", NBER Working Paper 15414.

Lambe, Geraldine (2007) "Public versus Private Equity Markets," The Banker, September, 2007

Leising, Matthew (2009) "ISDA Hires Rosen to Fight Obama OTC Derivatives Plan" Bloomberg, July 10th, 2009.

Levine, Ross (2005) "Finance and Growth: Theory and Evidence,"in *Handbook of Economic Growth*, Philippe Aghion and Steven Durlauf (eds.)

Morgenson, Gretchen (2010) "Fair Game: It's not over until it's in the Rules" The New York Times, August 29th, 2010.

Murphy, Kevin, Andrei Shleifer and Robert Vishny (1991) "The Allocation of Talent: Implications for Growth,"*Quarterly Journal of Economics*, Vol. 90, 630-49.

Philippon, Thomas (2008) "The Evolution of the US Financial Industry from 1860 to 2007: Theory and Evidence,"NYU Stern Working Paper.

Philippon, Thomas and Ariell Reshef (2008), "Wages and Human Capital in the US Financial Industry: 1926-2006," NYU Stern Working Paper, December.

Rothschild, Michael and Joseph Stiglitz (1976), Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information,"*Quarterly Journal of Economics*, Vol. 106, pp. 503-530.

Santos, Tano and Jose A. Scheinkman (2001) "Competition Among Exchanges," *Quarterly Journal of Economics,* August, 116 (3), pp. 1027-1061.

Scannell, Kara (2009) "Big Companies Go to Washington to Fight Regulations on Fancy Derivatives," The Wall Street Journal, July 10th, 2009.

Story, Louise (2010) "A Secretive Banking Elite Rules Trading in Derivatives" The New York Times, December 11th, 2010.

Sjostrom, William K. (2008) "The Birth of Rule 144A Equity Offerings," UCLA Law Journal, 56, pp. 409-449.

Tang, Vicki Wei (2007) "Economic Consequences of Mandatory Disclosure Regulation: Evidence from Rule 144A Equity Private Placements," McDonough School of Business Working Paper, Georgetown University.

Testy, Kellye Y. (1990) "The Capital Markets in Transition: A Response to New SEC Rule 144A," *Indiana Law Journal,* vol. 66, pp. 233-271.

Tett, Gillian (2010) "Calls for radical rethink of derivatives body," Financial Times, 26th of August, 2010.

Vayanos, Dimitri and Tan Wang (2007) "Search and Endogenous Concentration of Liquidity in Asset Markets," *Journal of Economic Theory,* vol. 136, pp. 66-104.

Vayanos, Dimitri and Pierre-Olivier Weill (2008) "A Search-Based Theory on the On-the-Run Phenomenon," *Journal of Finance*, vol. 63, pp. 361-1398.

Zingales, Luigi (2009) "The Future of Securities Regulation," *Journal of Accounting Research*, vol. 47, no. 2, May. pp. 391-425.

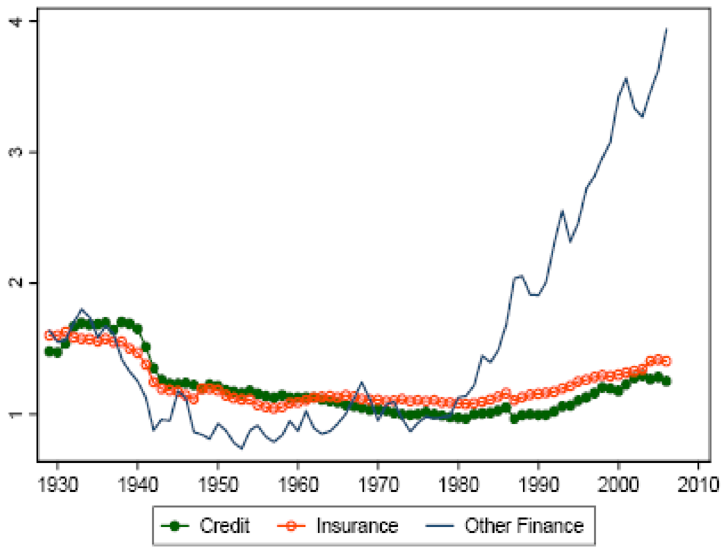

Figure 1: **Wages in finance relative to non farm private sector (Philippon and A. Reshef, 2008.)**

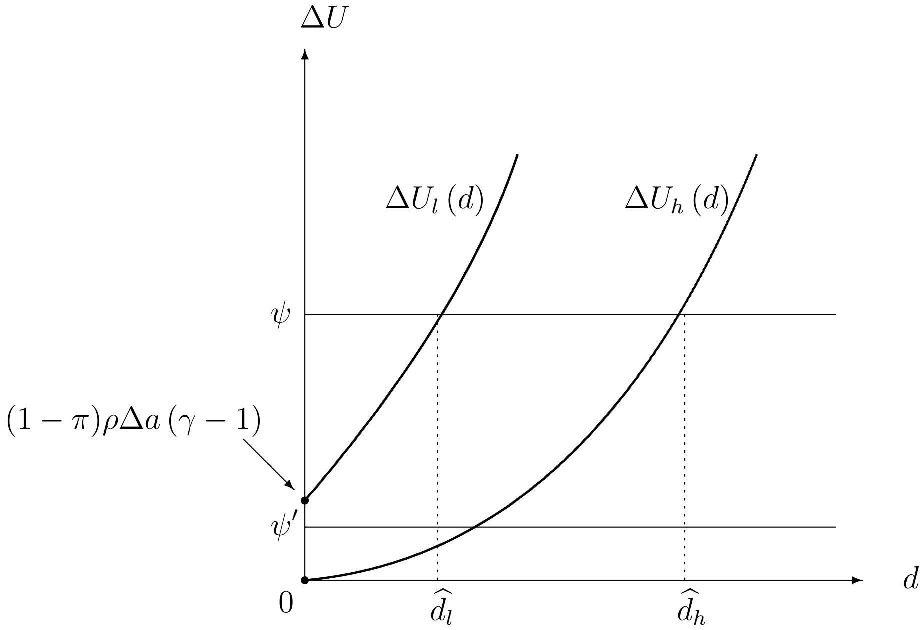

Figure 2: **Incentive compatibility.**

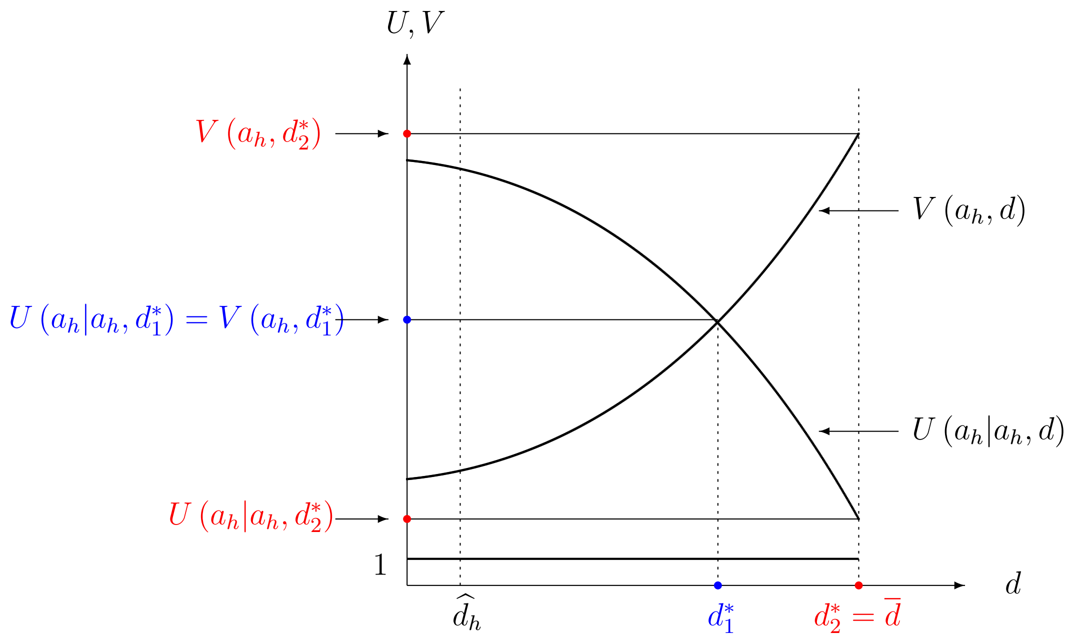

Figure 3: **High effort equilibria.**

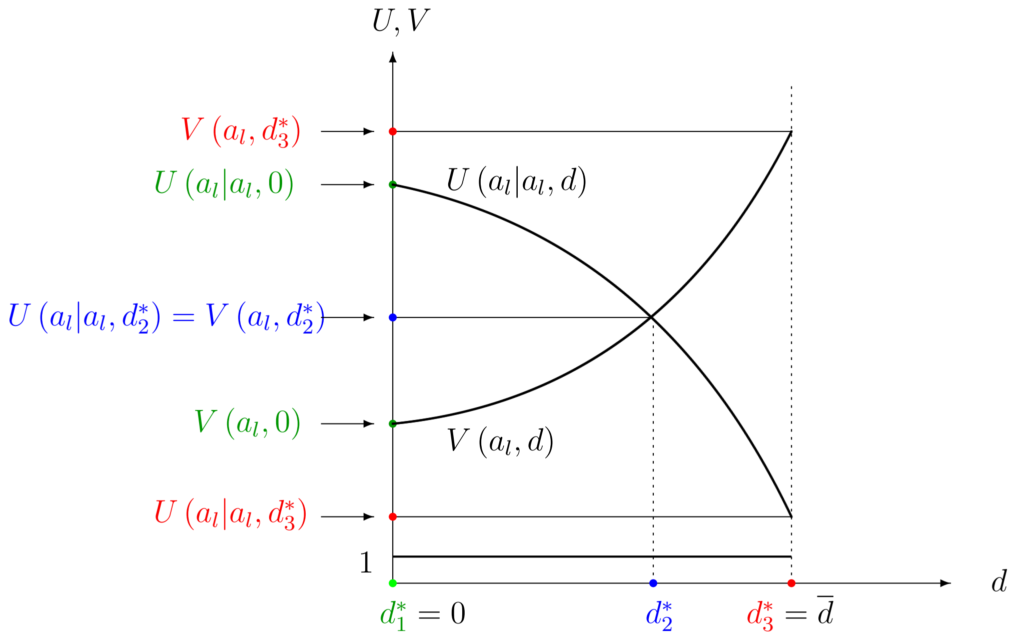

Figure 4: **Low effort equilibria.**

APPENDIX

Proof of Lemma 1 Consider first an impatient entrepreneur. By selling his asset in the organized market he is able to obtain at least p, which is higher than the maximum amount ρ he can borrow against the asset. Therefore, an impatient entrepreneur strictly prefers to sell his assets than to borrow. As for a patient entrepreneur, since he strictly prefers to consume in period 2 he cannot gain by borrowing and consuming in period 1. He also cannot gain (strictly) from borrowing and investing the proceeds from the loan in either the organized or OTC markets. A patient entrepreneur is no different as an investor than an uninformed type 1 agent, and therefore earns the same zero net returns in equilibrium as type 1 agents. Finally, consider an impatient dealer. This dealer is always better off consuming his endowment: Purchasing the asset, either in the OTC market or in the exchange, and borrowing against it can never be optimal since in both markets prices exceed ρ, the maximum amount he is able to borrow. $\square$

Proof of Lemma 2**.** A best response for a patient entrepreneur, who puts his asset up for sale in the OTC market is to always reject an offer from a dealer. Indeed, dealers only offer to buy good assets for a price $p^d < \rho\gamma$. The patient entrepreneur is then strictly better off holding on to an asset that has been identified as high quality by the dealer. If the asset that has been put up for sale does not generate an offer from an informed dealer, then the entrepreneur has the same uninformed value for the asset as type 1 agents. He is therefore indifferent between selling and not selling the asset at price p in the organized market. $\square$

In order to prove Proposition 3 it is useful first to prove Proposition 4.

Proof of Proposition 4. Note that

$$\frac{\partial m}{\partial d} = \frac{(1-\pi)}{\pi a\,(1-d)^2} > 0 \qquad \text{and} \qquad \frac{\partial^2 m}{\partial d^2} = \left(\frac{2}{1-d}\right)\frac{\partial m}{\partial d} > 0, \tag{25}$$

and

$$\frac{\partial p}{\partial d} = \left[\frac{a\rho\,(1-a)\,(1-\gamma)}{\left[a\,(1-m)+(1-a)\right]^2}\right]\frac{\partial m}{\partial d} < 0 \quad \text{as} \quad \gamma > 1. \tag{26}$$

Finally,

$$\frac{\partial^2 p}{\partial d^2} = \frac{a\rho\,(1-a)\,(1-\gamma)}{\left[a\,(1-m)+(1-a)\right]^2}\left[\frac{\partial^2 m}{\partial d^2} + \frac{2}{a\,(1-m)+(1-a)}\left(\frac{\partial m}{\partial d}\right)^2\right] < 0. \tag{27}$$

Expressions (25), (26), and (27) are used throughout. $\square$

Proof of Proposition 3. From (12),

$$\frac{\partial V}{\partial d} = -\,(1-\pi)\,(1-\kappa)\,\frac{\partial p}{\partial d} > 0,$$

and

$$\frac{\partial^2 V}{\partial d^2} = -(1-\pi)(1-\kappa)\frac{\partial^2 p}{\partial d^2} > 0,$$

given (26) and (27), which establishes (b).

As for the utility of the entrepreneur, (9), note that

$$\frac{\partial U}{\partial d} = \pi \frac{\partial p}{\partial d} + a\pi\kappa \left[\frac{\partial m}{\partial d}(\gamma\rho - p) - m\frac{\partial p}{\partial d}\right].$$

It can be shown that

$$\gamma\rho - p = \gamma\rho - \frac{a(1-m)\gamma + (1-a)}{a(1-m)+(1-a)}\rho = -\left(\frac{a(1-m)+(1-a)}{a}\right)\frac{\partial p/\partial d}{\partial m/\partial d},$$

and hence

$$\frac{\partial m}{\partial d}(\gamma\rho - p) - m\left(\frac{\partial p}{\partial d}\right) = -\left(\frac{a(1-m)+(1-a)}{a}\right)\frac{\partial p}{\partial d} - m\frac{\partial p}{\partial d} = -\frac{\partial p/\partial d}{a},$$

and thus we can write

$$\frac{\partial U}{\partial d} = \pi(1-\kappa)\frac{\partial p}{\partial d} < 0.$$

Finally,

$$\frac{\partial^2 U}{\partial d^2} = \pi(1-\kappa)\frac{\partial^2 p}{\partial d^2} < 0,$$

which proves (a). □

To prove Proposition 5 we first have to derive the utility of the entrepreneur under a deviation.

Proposition A. (a) Assume that the candidate action in equilibrium is $a^* = a_h$ then the utility of the entrepreneur who deviates and chooses instead to exercise action a_l is

$$U_{hl}(d) = p_h(d) + a_l m_h(d)(\gamma\rho - p_h(d))(\pi\kappa + (1-\pi)). \tag{28}$$

(b) Assume that the candidate action in equilibrium is $a^* = a_l$ then the utility of the entrepreneur who deviates and chooses instead to exercise action a_h is

$$U_{lh}(d) = -\psi + \pi[p_l(d) + a_h m_l(d)\kappa(\gamma\omega\rho - p_l(d))] + (1-\pi)\omega\rho[1 + a_h(\gamma-1)] \tag{29}$$

Proof. (a) The key is to show that if the entrepreneur deviates and instead exercises the low effort, then even in the absence of a liquidity shock he prefers to sell. For this define the following notation

$$U^{\text{sell}}(a_l|a_h, d, \text{no-liq.}) \qquad \text{and} \qquad U^{\text{no-sell}}(a_l|a_h, d, \text{no-liq.}), \tag{30}$$

the utility of the entrepreneur entering date 1 (that is, before being hit with bids (or no bids) by dealers) who (i) deviated from the high effort to implement the low effort at $t = 0$, (ii) does not suffer a liquidity shock at $t = 1$ and (iii) decides to sell and not sell, respectively, as a function of the measure of dealers, d. We want to show that

$$U^{\text{sell}}\left(a_l|a_h, d, \text{no-liq.}\right) \geq U^{\text{no-sell}}\left(a_l|a_h, d, \text{no-liq.}\right).$$

First, notice that

$$\begin{aligned} U^{\text{sell}}\left(a_l|a_h, d, \text{no-liq.}\right) &= a_l m_h(d)\gamma\rho + (1 - a_l m_h(d))p_h(d) \\ &= p_h\left(d\right) + a_l m_h\left(d\right)\left(\gamma\rho - p_h(d)\right), \end{aligned} \tag{31}$$

where the functions $p_h\left(d\right)$ and $m_h\left(d\right)$ were given by (10) and (11), respectively, when $a = a_h$. The first term in (31) is the payoff, conditional on having a good project and receiving a bid from a dealer, and event with probability $a_l m_h(d)$, in which case the entrepreneur rejects the bid and carries the project to maturity and obtains, $\gamma\rho$, as recall that he is not subject to the liquidity shock. The second term is the payoff when he does not receive a bid but sells anyway. Since $p_h = p_h(m(d))$ we may consider the function

$$f(m) = p_h\left(m\right) + a_l m\left(\gamma\rho - p_h(m)\right).$$

Notice that

$$U^{\text{no-sell}}\left(a_l|a_h, d, \text{no-liq.}\right) = \rho\left[1 + a_l\left(\gamma - 1\right)\right] = f(1). \tag{32}$$

Further,

$$\frac{\partial f}{\partial m} = \left(\frac{\rho\left(1 - a_h\right)\left(1 - \gamma\right)}{1 - a_h m}\right)\left(a_h\left(\frac{1 - a_l m}{1 - a_h m}\right) - a_l\right) < 0$$

Thus for every m, $f(m) \leq U^{\text{no-sell}}\left(a_l|a_h, d, \text{no-liq.}\right)$ establishing that

$$U^{\text{sell}}\left(a_l|a_h, d, \text{no-liq.}\right) \geq U^{\text{no-sell}}\left(a_l|a_h, d, \text{no-liq.}\right) \quad \text{for all} \quad d,$$

and hence

$$U_{hl}\left(d\right) = \pi\left[p_h(d) + a_l m_h(d)\kappa\left(\gamma\rho - p_h(d)\right)\right] + (1 - \pi)U^{\text{sell}}\left(a_l|a_h, d, \text{no-liq.}\right), \tag{33}$$

which after some manipulations yields (28).

(b) We show that

$$U^{\text{sell}}\left(a_h|a_l, d, \text{no-liq.}\right) \leq U^{\text{no-sell}}\left(a_h|a_l, d, \text{no-liq.}\right) \quad \text{for all} \quad d. \tag{34}$$

First notice that,

$$U^{\text{no-sell}}\left(a_h|a_l, d, \text{no-liq.}\right) = \rho\left[1 + a_h\left(\gamma - 1\right)\right].$$

Second, notice that

$$U^{\text{sell}}\left(a_h|a_l,d,\text{no-liq.}\right)=p_l(d)+a_hm_l\left(d\right)\left(\gamma\rho-p_l\left(d\right)\right),$$

and, as before, define

$$g\left(m\right)=p_l(m)+a_hm\left(\gamma\rho-p_l\left(m\right)\right).$$

Notice that

$$U^{\text{no-sell}}\left(a_h|a_l,d,\text{no-liq.}\right)=\rho\left[1+a_h\left(\gamma-1\right)\right]=g(1).$$

Finally, we can show that

$$\frac{\partial g}{\partial m}=\left(\frac{\rho\left(1-a_l\right)\left(1-\gamma\right)}{1-a_lm}\right)\left(a_l\left(\frac{1-a_hm}{1-a_lm}\right)-a_h\right)>0.$$

Thus for every m, $g(m)\leq U^{\text{no-sell}}\left(a_h|a_l,d,\text{no-liq.}\right)$, establishing (34). Thus

$$U\left(a_h|a_l,d\right)=-\psi+\pi\left[p_l(d)+a_hm_l(d)\kappa\left(\gamma\omega\rho-p_l(d)\right)\right]+(1-\pi)U^{\text{no-sell}}\left(a_h|a_l,d,\text{no-liq.}\right). \quad (35)$$

Trivial manipulations of (35) yield (29). □

Proof of Proposition 5**.** (a) It can be shown that

$$\frac{\partial\Delta U_h}{\partial d}=-\frac{\partial p_h}{\partial d}\frac{\Delta a}{a_h}\left[\pi\kappa+(1-\pi)\right]>0,$$

by Proposition 4. Similarly notice that

$$\frac{\partial\Delta U_l}{\partial d}=-\frac{\partial p_l}{\partial d}\frac{\Delta a}{a_l}\pi\kappa>0.$$

(b)

$$\begin{aligned}
\Delta U_h(d) &= \pi\Delta am_h(d)\kappa\left(\gamma\rho-p_h(d)\right)\\
&+ (1-\pi)\left[\rho\left(1+a_h\left(\gamma-1\right)\right)-\left(p_h(d)+a_lm_h(d)\left(\gamma\rho-p_h(d)\right)\right)\right]\\
&< \pi\Delta am_h(d)\kappa\left(\gamma\rho-p_h(d)\right)\\
&+ (1-\pi)\left[\rho\left(1+a_h\left(\gamma-1\right)\right)-\rho\left[1+a_l\left(\gamma-1\right)\right]\right]\\
&= \pi\Delta am_h(d)\kappa\left(\gamma\rho-p_h(d)\right)+(1-\pi)\rho\Delta a\left(\gamma-1\right)\\
&< \pi\Delta am_l(d)\kappa\left(\gamma\rho-p_l(d)\right)+(1-\pi)\rho\Delta a\left(\gamma-1\right)\\
&= \Delta U_l(d),
\end{aligned}$$

as

$$m_l(d)>m_h(d) \qquad \text{and} \qquad p_l(d)<p_h(d),$$

by Proposition 4. □

Proof of Proposition 6. (a) If $\Delta U_h(d) \geq \psi$ then by Proposition 5, $\Delta U_l(d) > \psi$. Thus $\widehat{d}_h \geq \widehat{d}_l$. (b) and (c) follow from the strict monotonicity of $\Delta U_l(d)$ and $\Delta U_h(d)$ and the fact that $\psi > 0$. □

Proof of Proposition 7. This follows immediately from the observation that all high effort equilibria have a measure of dealers $d \in [\widehat{d}_h, \overline{d}]$. If for a vector of parameters $\mathcal{A}$ there is an equilibrium for which $d_h^* = \widehat{d}_h$, consider a vector $\mathcal{A}'$ with entries identical to $\mathcal{A}$ but a smaller a_l. Then the optimality of high effort is maintained and the function $\Delta U_h(d)$ shifts up. Hence $\widehat{d}'_h < \widehat{d}_h$. However the original high effort equilibrium is still an equilibrium since all parameters where kept except for a smaller a_l. Further, any high effort equilibrium such that $d_h^* > \widehat{d}'_h$ is inefficient □

Proof of Proposition 8. This follows immediately from the monotonicity of the expression in (23) . □