Is the financial sector too big?

Patrick Bolton
Columbia University

Tano Santos
Columbia University

Jose A. Scheinkman
Princeton University

September 8, 2010

Abstract

We propose an equilibrium occupational choice model, where agents can choose to work in the real sector (become entrepreneurs) or in the financial sector (become dealers). Agents incur costs to become informed dealers and develop skills in valuing assets up for trade. The financial sector comprises an organized competitive exchange, where uninformed agents trade and an over-the-counter (OTC) market, where informed dealers are ready to offer attractive terms for the most valuable assets entrepreneurs put up for sale. Thanks to their information advantage and valuation skills dealers are able to provide incentives to entrepreneurs to originate good assets. However, the opaqueness of the OTC market allows dealers to extract informational rents from entrepreneurs. Trade in the OTC market also imposes a negative externality on the organized exchange, where only the less valuable assets end up for trade. We show that in equilibrium the dealers' informational rents in the OTC market are too large and attact too much talent into the financial industry.

1 Introduction

What does the financial industry add to the real economy? What is the optimal size of the financial sector relative to the economy? 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? [*Wall Street Journal*, December 14, 2009]

The issue is not so much whether the financial industry helps channel household savings to fund real investments, or whether it is a provider of liquidity and helps investors diversify risk. The fact that the financial industry performs these basic functions is all well understood. Rather, the issue is whether the financial industry extracts excessively *high rents* from these activities and whether it attracts too much *young talent*. In this paper we propose an equilibrium model with endogenous occupational choice between the financial and the real sector, in which the financial industry does indeed extract excessive informational rents and attracts too much talent.

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. That is, if lending and capital provision by the financial industry were to grow, output and welfare would also grow. In other words, most of these models conclude that due to asymmetries of information, and incentive or contract enforceability constraints, there is underinvestment in equilibrium and *financial underdevelopment*. These models provide theoretical underpinnings to the empirical findings from cross-country regressions that financial development leads growth.

In contrast to this literature, our model mainly emphasizes the fifth function of the financial industry 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 *organized exchanges* and over-the-counter (OTC) markets.

As is illustrated in Figure 1, which plots the evolution of wages in the U.S. banking, insurance and 'other finance' sectors from the great depression onwards, the key growth in remuneration in the financial industry at large has taken place in investment banking, derivatives trading and OTC markets. Thus, in order to understand whether the financial industry extracts too high rents and as a result attracts too much talent, one needs to focus on why remuneration in these markets is so high.

In our model, secondary market trading requires information about underlying asset quality and valuation skills. When an entrepreneur is looking to sell his firm in the secondary market the buyer must be able to determine the value of the firm that is up for sale. This is where the *young talent* employed in the financial industry manifests itself. *Informed dealers* in the OTC market are able to determine the value of assets for sale and can offer to buy the most valuable assets from entrepreneurs. 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 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. Importantly, in this market price offers of dealers and negotiated transactions are not disclosed. This is in contrast to the organized market where all quotes and transactions are disclosed. As a result of the scarcity of informed dealers and *opacity* of the OTC market, informed dealers are able to extract an *informational rent* from the entrepreneurs selling the most valuable assets to them. Indeed, 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.

In other words, as in Rothschild and Stiglitz (1976), informed dealers in the OTC market are able to *cream-skim* the best assets and thus extract an informational rent. This 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.

This is why in our model rent-extraction by informed dealers actually increases as the OTC market expands. This is also the key mechanism which: *i)* gives rise to an excessively large financial industry in equilibrium; *ii)* is why there is excessive information rent-extraction in the OTC market; and, *iii)* is why the financial sector–in particular the OTC market–attracts too much talent. All the human capital invested in young talent to train an informed dealer in the end mainly serves to extract informational rents rather than create social surplus by allowing entrepreneurs who originate valuable assets to realize a fair value for those assets when they sell them in secondary markets. 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 informal 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 cannot be reflected in a market price, as each transaction is an undisclosed bilateral deal between the dealer and the seller of the asset. Therefore, when more dealers compete in the OTC market, this does not result in more information being transmitted through prices. On the contrary, more competition by informed dealers simply results in more cream-skimming and more information rent extraction. Our highly stylized model of the financial industry can be seen as an *allegory* of a general phenomenon in the financial industry, where informed parties have an incentive to trade and remove themselves from organized markets. This is not just true for derivatives and swaps, which are mostly traded in OTC markets, but also for secondary stock markets, where trading by institutions often takes place in an informal 'upstairs market' or more recently in so called 'dark pools'. Similarly, for primary markets, private equity funding or private placements have the same cannibalizing effect on organized exchanges, of removing the better and hardest to value assets from uninformed investors' reach.

Our paper contributes to a small literature on the optimal size of 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 to show that in a two-sector model there may be inefficient equilibrium occupational outcomes, where too much talent enters one market since the marginal private returns from talent 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, Lagos, Rocheteau and Weill (2009) propose a model of the OTC market that has common elements to ours. However, their focus is on the liquidity of this market and they do not address issues of cream-skimming or occupational choice.

The paper is organized as follows: Section 2 outlines the model. Section 3 analyzes entrepreneurs' moral hazard in origination problem. Section 4 considers agents' ex-ante occupational choice problem between the financial and real sectors and characterizes the general equilibrium. Section 5 examines the efficiency of equilibrium occupational choices. Finally, Section 6 concludes.

2 The model

We consider a competitive economy divided into two sectors: a real, productive, sector and a financial sector.

2.1 Agents.

The economy is comprised of a continuum of three-period lived, risk-neutral, agents who can be of two different types. Type 1 agents 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 are the *active population.* Each type 2 agent can 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 to which they are committed to for the remainder of their life.

The core of our model centers on the interaction between the real and financial sectors. On the one hand, these two sectors complement each other, as the real sector can be an efficient source of assets only to the extent that the financial sector provides funding, liquidity, and valuation services for the assets originated in the real sector. On the other hand, these two sectors are also compete for scare human capital, the type 2 agents in our model.

We simplify the model without much loss in generality by assuming that all type 2 agents start in period 1 with the same unit endowment, $\omega = 1$, have the same preferences over consumption, face the same idiosyncratic liquidity shocks, and are equally able entrepreneurs. 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 the increasing order of the costs they face of acquiring the human capital to become well informed dealers: $\varphi(d)$. We then assume that $\varphi(d)$ is non-decreasing and that in addition there is an agent $\overline{d} < 1$ that effectively bounds the maximum size of the financial sector:

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

In all other respects, type 2 agents are identical:

1. They face the same *i.i.d.* liquidity shocks[1] and value consumption only in period 1 with probability π and only in period 2 with probability $(1 - \pi)$. Their liquidity preferences, whether they choose to become entrepreneurs or dealers, are thus 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 $(\delta = 1) = \pi$.[2]

2. 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 $0 < a_l < a_h \leq 1$.[3] If he chooses $a = a_l$ then his *effort cost* $\psi(a_l)$ is normalized to zero, but he is then only

[1]We could have instead attached the liquidity shock to the activity rather than to the agent and thus allowed for different liquidity shocks for entrepreneurs and dealers. The analysis would have been conceptually identical to the present one though more notationally intensive. To present the simplest case we opt to assume that all type 2 agents face the same liquidity shocks.

[2]To keep our notation as streamlined as possible we use the lower case $u(\cdot)$ for the utility function of type 1 agents and the upper case $U(\cdot)$ for the utility function of type 2 agents.

[3]In the robustness subsection we relax the assumption that $a_l > 0$ and allow for the possibility that $a_l = 0$.

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. Note that patient entrepreneurs have no output in period 1 that they could trade with impatient entrepreneurs.

3. 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 non-pecuniary 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, to purchase assets for sale by impatient entrepreneurs. If they learn that they are impatient they simply consume their unit endowment. For simplicity in what follows we assume that patient dealers can only acquire one unit of the asset at date 1, an assumption to which we return below and that we also discuss at length in in the robustness section.

2.2 Financial Markets

A central 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 on 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 go to the organized exchange and sell his asset for the competitive equilibrium price p; 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. However, just as in Grossman and Stiglitz (1980), dealers' information must be in scarce supply in equilibrium, as dealers must be compensated for their cost $\varphi(d)$ or acquiring their valuation skills. As will become clear below, this means not only that dealers only purchase high quality assets worth $\gamma\rho$ in equilibrium, but also that not all entrepreneurs with high quality assets will be able to sell to a dealer.

Thus, 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 have no choice but to turn to the organized market in which they can sell their asset for price p.

We assume that the probability m is simply given by the ratio of the total mass of patient dealers $d(1-\pi)$ divided by 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. 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 a bilateral bargaining game (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*, where the dealer has bargaining power $(1-\kappa)$ and the entrepreneur has bargaining power κ (see Nash, 1950, 1953).[4] 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

[4] For a similar approach to modeling negotiations in OTC markets between dealers and clients see Lagos, Rocheteau, and Weill (2009).

Binmore, Rubinstein and Wolinsky, 1986).[5]

The price p^d may be higher than 1, the dealer's endowment. In that case the dealer needs to borrow the difference $(p^d - 1)$ against the asset to be acquired.[6] As long as this difference does not exceed ρ, the dealer will not be financially constrained. For simplicity, we shall 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.[7]

Consider next the organized exchange. This is a competitive market in which all low quality assets $(1-a)(1-d)\pi$ are traded as well as a fraction $(1-m)$ of high quality assets $a(1-d)\pi$. The buyers of assets are uninformed rentiers, who are unable to distinguish high quality from low quality assets. Entrepreneurs, themselves do not know the true underlying quality of their assets. All rentiers and entrepreneurs can ascertain is the expected value of their asset, conditional on being turned down by dealers in the OTC market:

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

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 $m = 1$.

2.3 Timing

To summarize, the timing in the model is as follows:

[5]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 $\kappa = \alpha$.

[6]To simplify the exposition we assume throughout that the dealer cannot acquire more than one project.

[7]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.

1. In period 0, type 2 agents choose between the occupations of entrepreneur or dealer based on which will yield a higher expected payoff. Dealers incur a personal cost $\varphi(d)$ of becoming informed dealers, and entreprepreneurs use their endowment to invest in a project and also choose an effort level $a \in \{a_l, a_h\}$. Let $\hat{d}$ be a type 2 agent with cost $\varphi(\hat{d})$ of becoming a dealer, then under our assumption that $\varphi(d)$ is non-decreasing in d, if agent $\hat{d}$ prefers to become a dealer then all agents $d \in [0, \hat{d}]$ also prefer to become dealers.

2. At the beginning of period 1 liquidity shocks are realized and type 2 agents learn whether they are patient or impatient to consume. At the same time the underlying value of the assets originated by entrepreneurs is determined.

3. All impatient dealers then consume their endowment, and all impatient entrepreneurs seek out a patient dealer to sell their asset to. All patient dealers eventually end up *matching* with an entrepreneur with a high quality asset. They negotiate a deal for that asset for a price $p^d = \kappa\gamma\rho + (1-\kappa)p$ and entrepreneurs go on to consume p^d. Patient dealers borrow from rentiers an amount $\max\left\{p^d - 1, 0\right\} < \rho$ against this asset.

4. The impatient entrepreneurs who do not match with a patient dealer, put their asset for sale in the organized exchange at price p given in equation (5) and consume p.

5. Type 1 agents (rentiers) are indifferent as to when they consume. Without loss of generality we adopt the convention that they consume all their endowment in period 2. That is, those rentiers who did not purchase any assets from entrepreneurs consume their endowment ω. Those who did purchase assets from entrepreneurs consume $\omega + (\gamma\rho - p)$ if they were lucky to end up with a high quality asset, or $\omega - (p + \rho)$ if they were unlucky and purchased a bad quality asset.

6. Patient type 2 agents strictly prefer to consume in period 2. Thus, patient dealers consume their net claim to period 2 output $\gamma\rho - (p^d - 1)$.

7. As for patient entrepreneurs, we will show that along the equilibrium path they hold on to the asset they originated in period 0 until maturity and then consume the asset's

output. Their expected consumption is then $\rho[a(\gamma-1)+1]$. In a symmetric equilibrium where all entrepreneurs choose $a = a_h$, we then need to check that the double deviation, where a single entrepreneur chooses $a = a_l$ in period 0 and sells his asset in period 1 even if he learns that he is patient is not profitable.

2.4 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 in a straightforward way, 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 for simplicity.

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. They are able to reward entrepreneurs for originating good assets and thereby they provide incentives to entrepreneurs to put in high effort to originate good assets. If it were not for these positive incentive effects, informed dealers would mostly play a parasitical role in our economy. They 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 non-pecuniary costs in acquiring information to become a dealer. We could also, or alternatively, have introduced heterogeneity in the costs of becoming an entrepreneur. Nothing substantive would be added by introducing this other form of heterogeneity. 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.

The reader may wonder why we introduce any ex ante heterogeneity among type 2 agents at all? It turns out that the greater generality of the model with ex-ante heterogeneous type 2 agents actually gives rise to an analytically more tractable model. Indeed, with ex-ante identical type 2 agents, all these agents would have to be indifferent between becoming dealers or entrepreneurs in equilibrium and supporting such an equilibrium would require that type 2 agents randomize their choice between their two occupations. Characterizing such a mixed-strategy equilibrium would be analytically more involved and would not give rise to a simpler analysis.[8]

As we have argued above, we shall restrict attention to parameter values for which the measure of patient dealer 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. 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 $p^d - 1 \leq \rho$. For this it is enough to assume that

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

[8]It would also not explain the observed high rents in the financial sector.

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

$$\rho\left[1 + a_l\left(\gamma - 1\right)\right] \geq 1.^{9} \tag{8}$$

2.5 Definition of equilibrium

A general equilibrium in our economy 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 must 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 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 last equilibrium is going to be much the focus of what follows as it is only in this case in which, as we will show, 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.

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.

[9] Importantly this does not mean that such an allocation would be an equilibrium.

3 Equilibrium payoffs and the moral hazard problem

There are basically two non-trivial optimization problems in our model. The first is the occupational choice problem of type 2 agents. The second is the effort choice problem of entrepreneurs in period 0. 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 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.[10]

3.1 Equilibrium borrowing and asset trading in period 1

Before characterizing the equilibrium occupational choices of type 2 agents in period 0 we begin by describing equilibrium play in period 1 in either equilibrium. In period 1, d^*, a^* and, $m(a^*, d^*)$ are given. For any (a^*, d^*) we are able to establish the following first lemma.

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

The most interesting result of the lemma is that impatient entrepreneurs are better off selling their assets than to borrow against their asset to finance their consumption. 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

[10] Recall that, as we have already argued in section 2.2, entrepreneurs in distress always prefer to exercise the (free) option of attempting to sell in the dealer market and only when they are not matched with an informed dealer to sell in the organized exchange.

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 *A patient entrepreneur, who follows the equilibrium action $a^* \in \{a_l, a_h\}$ in period 0 (weakly) prefers not to put up his asset for sale in period 1.*

Thus in our framework the equilibrium plays in a standard way: Entrepreneurs subject to liquidity shocks sell whereas patient carry their projects to date 2.

3.2 Equilibrium Payoffs in period 0

We are now in a position to determine equilibrium payoffs for dealers and entrepreneurs in period 0, that is the expected payoff of, for example, the entrepreneur when his chosen action is the equilibrium one. Given that our focus is on the occupational choice of type 2 agents, we characterize these expected payoff functions as a function of the measure of dealers, d. Recall that we focus on incentive compatible symmetric equilibria and thus all entrepreneurs are identical; only dealers differ as they have potentially different costs of acquiring information. Let $U(a|a', d)$ 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 $V\left(\widetilde{d}|a', d\right)$ 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 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.

Notice that all entrepreneurs have identical expected utility. This is not the case with dealers as we have assumed that the only source of ex-ante cross sectional heterogeneity is in the costs of acquiring information. Let $V\left(\widetilde{d}|a^*,d\right)$ 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. The payoff functions are depicted in Figure 2.

Proposition 3 *(a) The utility of the 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: The larger the measure of dealers the more likely the fewer remaining entrepreneurs are to be matched with some dealer. (b) is at the heart of our paper. As the number of dealers increase 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 payoff function. 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.[11] This results in an increasing expected payoff for the dealers as a function of d. 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 it 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 and thus what dealers are willing to bid for the asset in the OTC market, which,

[11] We ignore for the time being the effect that additional competition amongst dealers may have on their rents, which as we show below, does not alter the main results of this paper.

of course hurts all entrepreneurs in distress, whether they get matched or not. 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, which is precisely because of the larger rents that dealers capture due to cream-skimming as d goes up. 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.

Finally we note another interesting implication of our model, which is that dealers also prefer dealing in market equilibria with low quality origination of assets. The reason is that, as shown in Proposition 4 (b), 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 dealer could induce bad asset origination they would do so.

The reader may have the impression that dealers serve no welfare enhancing purpose in our framework, but this is not the case. As we will show next a strictly positive measure of dealers is needed to support the high effort. If it were not for the positive incentive effects of cream skimming in OTC markets on entrepreneurs, informed dealers would mostly play a parasitical role in our economy. They would enrich themselves by helping entrepreneurs with good assets get a better price, but they would not create any net social surplus. We turn next then to the moral hazard problems at $t = 0$ and the role of OTC markets in relieving these moral hazard problems.

3.3 Entrepreneur Moral Hazard

Consider an entrepreneur in period 0 who has made his physical investment and is deciding whether to choose the low effort a_l or the high effort a_h. This entrepreneur is looking forward to what may happen in periods 1 and 2 when facing this decision, and this in turn depends on what all other entrepreneurs are rationally expected to do. That is, the market outcome in period 1 depends on whether financial markets expect entrepreneurs to choose action a_l or a_h.

A necessary condition for any symmetric equilibrium then is that it is *incentive compatible* for entrepreneurs to choose the equilibrium effort: They must (weakly) prefer to choose the equilibrium action than to deviate to the other action:

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

Throughout we write $U_h(d)$, as in (14) below, for the equilibrium expected payoff of the entrepreneur along the high effort equilibrium path as a function of d, where $p_h^d(d)$, $p_h(d)$ and $m_h(d)$ refer to the prices and matching probabilities. We denote by $U_{hl}(d)$ the utility of the entrepreneur that deviates and instead implements action a_l instead of a_h, that is,

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

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:

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

Suppose now that an entrepreneur chooses to *deviate* in period 0 by choosing the low effort a_l. As we show below in the appendix in the proof of Proposition 9, in this case it is optimal for such an 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 agents 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, even when there is substantial cream skimming, as the rest of the entrepreneurs implement the high effort. Thus when selling in the uninformed exchange, the shirking entrepreneur benefits by hiding behind these better projects to obtain a subsidy. More formally, Proposition 9 shows that the payoff of an entrepreneur that deviates to the low effort when the measure of dealers is given by d is,

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

If the measure of dealers is d then the 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*, that is net of high effort costs 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

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

that is, that the expected monetary payoff of adhering to the high effort rather than deviating to the low effort compensates for the cost of the high effort provision.

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 then:

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

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 9 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 in period 1. The reason is that now his private valuation is higher than the average quality of the assets in the exchange. His expected payoff under the deviation is given by:

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

which is also derived in Proposition 9 in the Appendix.

Incentive compatibility in the low effort equilibrium when the measure of dealers is d again requires that $U_l\left(d\right) \geq U_{lh}\left(d\right)$, or if we define $\Delta U_l\left(d\right)$ as the difference in expected *monetary payoffs*, that is, net of 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:[12]

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

then incentive compatibility requires that

$$\Delta U_l\left(d\right) \leq \psi, \tag{19}$$

that is, that the expected monetary payoff from deviating from the low effort to the high effort does not compensate for the cost of high effort provision ψ.

A critical step now is the characterization of the functions $\Delta U_h\left(d\right)$ and $\Delta U_l\left(d\right)$, which we do in the next proposition.

Proposition 5 *(a) $\Delta U_h(d)$ and $\Delta U_l(d)$ are both 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 $\Delta U_l(d)$ and are shown in Figure 3. The reason why these functions are increasing functions of 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

[12]Notice thus the slight asymmetry in the definition of ΔU_{hl} and ΔU_{lh}. To reiterate, ΔU_{hl} is the difference between the expected monetary payoff that accrues to the agent when he implements the candidate equilibrium high effort action, a_h, and the one he could obtain were he to instead deviate and implement a_l. ΔU_{lh} is the difference between the expected monetary payoff that accrues to the entrepreneur who deviates and implements a_h, instead of the candidate low equilibrium low effort a_l, and the expected monetary effort that he obtains when he adheres to the candidate low effort action. We opt for this definition to offer the intuition in terms of incentives to implement action a_h, whether from the candidate high or the low effort equilibrium.

in the event that he has a good asset. Therefore his incremental payoff from deviating is larger. As 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.

As for (b) the reason for the fact that $\Delta U_h(d) < \Delta U_l(d)$ has to do with the different out-of-equilibrium behavior of the entrepreneur when he deviates from the high effort equilibrium and the low effort one. Effectively when entrepreneurs are implementing the high effort the deviant agent has "more options" than when they are implementing the low effort. The reason is that the 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, thus receiving a subsidy. This is not the case in the low effort equilibrium as then the deviant entrepreneur implements a_h and were he to sell 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 $\widehat{d_l}$ respectively by the following equations

$$\Delta U_h(\widehat{d_h}) = \psi \qquad \text{and} \qquad \Delta U_l(\widehat{d_l}) = \psi, \tag{20}$$

we are able to establish our first major characterization of equilibrium occupational choice in period 0 in the following proposition.

Proposition 6 *(a) $\widehat{d_l} < \widehat{d_h}$. (b) A low-origination-effort equilibrium can only be supported for $d \in [0, \widehat{d_l}]$ and in particular no low effort equilibrium exists when $\psi < (1-\pi)\rho\Delta a\,(\gamma - 1)$. (c) A high-origination-effort equilibrium can only be supported for $d \in [\widehat{d_h}, 1]$ and $\widehat{d_h}$ is such that $\widehat{d_h} > 0$; and (d) there is no equilibrium with $d \in (\widehat{d_l}, \widehat{d_h})$.*

Proposition 6 is key in establishing the main results of the paper and merits emphasizing some of its implications. Figure 4 is simply Figure 3 where we have added two possible costs of exercising the high and the low effort, ψ and ψ'. First notice that the high effort is *never* incentive compatible in the absence of a financial sector, that is, when $d = 0$. 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. This result is somewhat surprising: 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 putative high effort equilibrium, the price of the asset in the uninformed exchange is very large when d is close to 0. There are two reasons for this. First in this case there is a large measure of entrepreneurs, $1 - d$, all exercising the high effort; second there is little cream skimming and thus 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. The deviation is profitable because when exercising the low effort the agent 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.[13] 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 agents prefer simply to save on effort costs and free ride on the large pool of entrepreneurs exercising the high effort. Notice as well that this results holds even when a_h is low and close to a_l precisely because in that case the benefits of adhering to the high effort over the low one are small.

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, independently of the measure of informed dealers. Indeed there is a sharp condition that is sufficient to rule the existence of a low effort equilibrium:

$$(1 - \pi)\rho\Delta a (\gamma - 1) > \psi, \tag{21}$$

which is the case when $\psi = \psi'$ in Figure 4. The argument is as follows. When entrepreneurs are playing the low effort, the price in the uninformed exchange is low. Thus the entrepreneur,

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

given that effort is not very costly, 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 is likely to be matched to an informed dealer in case of a liquidity shock as in this case there are not many entrepreneurs with high quality projects due to their low effort provision. These two effects are increasing in Δa. Indeed as can be seen in Figure 4 and in (21) the range of ψs for which a low effort equilibrium does not exist is increasing in Δa.

The next section shows the main results of the paper, which are immediate corollaries of the previous results.

4 The equilibrium size of the financial sector and welfare

4.1 Equilibrium Size of the Financial Sector

We now turn to a key question we are interested in: what is the equilibrium size of the financial sector? In our model this question boils down to determining the equilibrium measure of dealers d^*. As we have already highlighted, there may be two types of equilibria, each with an associated size of the OTC market. One type of equilibrium is the low-origination-effort equilibrium, in which all entrepreneurs choose $a = a_l$. As we saw in Proposition 6, this equilibrium can only be supported when $d \leq \widehat{d}_l$. The other type of equilibrium is the high-origination-effort equilibrium, in which all entrepreneurs choose $a = a_h$, and 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.

To offer a sharp characterization of the result it is useful, but not necessary, to simplify slightly the model and assume the following functional form for the costs of acquiring the information:

$$\varphi(d) = \overline{\varphi} \quad \text{for} \quad d \leq \overline{d} \qquad \text{and} \qquad \varphi(d) = +\infty \quad \text{for} \quad d > \overline{d}, \tag{22}$$

and thus the maximum possible size of the financial sector is given by $\overline{d}$. Under (22) all dealers are identical and thus when plotting the expected payoff function of any of them one also plots that of the *marginal* dealer who determines the size of the OTC market and thus we write $V(a, d)$ instead of $V\left(\widetilde{d}|a, d\right)$ for simplicity.

It is relatively simple to construct examples of economies for which there is no equilibria 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 the three possible cases: One of 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 (22), d^* has to be such that

$$U(a^*|a^*, d^*) \geq V(d|a^*, d^*) \qquad \text{for} \qquad d \geq d^*$$
$$U(a^*|a^*, d^*) < V(d|a^*, d^*) \qquad \text{for} \qquad d < d^*.$$

4.1.1 High effort equilibria

Consider the following parameter values

$$a_h = .75 \quad a_l = .55 \quad \gamma = 1.5 \quad \rho = .8 \quad \kappa = .25 \quad \pi = .5. \tag{23}$$

We also choose

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

where $\overline{d}$ was defined in (22).[14] Simple numerical calculations show that $m \leq 1$ for $d \leq .4286$ (see expression (6)) and thus, given that $\overline{d} = .35$, the matching probability is less than one is never a biding constraint. There is no low effort allocation that is incentive compatible in this example as

$$(1-\pi)\rho\Delta a(\gamma - 1) > \psi,$$

which implies $\Delta U_l(d) > \psi$ for all $d \geq 0$ (see Figure 4.)

As for high effort allocations these are incentive compatible as long as $d \geq \widehat{d}_h = .0536$, where $\widehat{d}_h$ was defined in (20). There are two high origination-effort equilibria and they are shown in Figure 5. First, there is an unstable equilibrium with $d_1^* = .3106$ in which all agents $d \leq \overline{d}$ are indifferent between becoming entrepreneurs or dealers. Second, there is a stable equilibrium with $d_2^* = \overline{d} = .35$, in which dealers are *strictly* better off as such than as entrepreneurs. Notice that all agents who can become dealers are dealers in equilibrium and thus our economy is at a corner.

[14] In this case (2) should be minimally modified to $\lim_{d \to \overline{d}^+} \varphi(d) = +\infty$.

Note that the price of assets in the OTC market in the unstable equilibrium is then $p^d(a_h, d_1^*) = 1.0180$, so that a dealer needs some leverage in order to finance the purchase of the asset. In contrast, in the stable equilibrium leverage is not needed as $p^d(a_h, d_2^*) = .9833$, which is less than their endowment.

4.1.2 Low effort equilibria

Consider now an example that generates only low effort equilibria. Take the parameters as in (23) and instead assume that

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

There are no high effort equilibria in this example as with this parameter specification $\Delta U_h(d) < \psi$ for $d \in [0, .15]$ (see Figure 4).

It can be numerically shown that all possible occupational choices yield incentive compatible low effort allocations as $\Delta U_l(d) < \psi$ for all $d \in [0, .15]$. As shown in Figure 6, there are then three (low origination-effort) equilibria. First there is an stable equilibrium where $d_1^* = 0$. Indeed, notice that when there are no dealers $U(a_l \mid a_l, 0) > V(0 \mid a_l, 0)$. Second, there is an unstable equilibrium with a measure of informed dealers $d_2^* = .0781$ with $U(a_l \mid a_l, .0781) = V(.0781 \mid a_l, .0781)$, that is, the marginal dealer is indifferent between becoming one or an entrepreneur. Finally, there is a stable equilibrium with $d_3^* = .15$ where $U(a_l \mid a_l, .15) < V(.15 \mid a_l, .15)$.

4.1.3 Coexistence of high and low effort equilibria

One can generate examples where there is both high and low effort equilibria. Consider for example the case where

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

and the rest of the parameters are as in (23) with the exception of $\kappa = .5$. In this case it can be shown that there are three equilibria, two that feature the low effort equilibria and one stable high effort equilibrium.

Start with the low effort equilibria. First, $\widehat{d}_l = .0545$ and in this region there are two equilibria. A stable one that features $d_1^* = 0$ as $U(a_l|a_l, 0) > V(0, a_l, 0)$ and an unstable one where type 2 agents are indifferent between becoming dealers or entrepreneurs and where the equilibrium measure of dealers is given by $d_2^* = .05$.

As for high effort allocations these are only incentive compatibles in the region $d \in \left[\widehat{d}_h, \overline{d}\right] = [.4020, .41]$. There is a candidate high effort equilibrium at which type 2 agents are indifferent between becoming dealers or entrepreneurs, at $d = .3596$ but is not incentive compatible. The allocation $(a_h, d_3^* = .41)$ is thus a stable high effort equilibrium. Moreover all three equilibria meet the participation constraint in that equilibrium utilities are above the reservation value of 1, which type 2 agents can always obtain by abstaining from becoming either dealers or entrepreneurs and simply carry their endowment forward.

4.2 Welfare: the inefficiently large size of the financial sector

4.2.1 The concept of constrained efficiency

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 only allow the planner to dictate the occupation of type 2 agents and we do not let the planner make any decisions based on the information obtained by informed dealers. The planner's problem in period 0 is then to pick the measure d of type 2 agents that maximizes ex-ante social surplus. If it is socially efficient to implement the low origination-effort a_l, then the efficient allocation consistent with that outcome, d_l^{ce}, is such that $d_l^{ce} \in [0, \hat{d}_l]$, as this is the region where the low effort equilibrium is incentive compatible. If instead it is socially efficient to implement the high origination-effort a_h, then the socially efficient allocation consistent with that outcome, d_h^{ce} must be such that $d_h^{ce} \in [\hat{d}_h, \overline{d}]$, where $\overline{d}$ is defined in (2).

4.2.2 The size of the financial sector and constrained efficiency

To establish as sharp a characterization as possible we focus first in situations where there is a role for the financial sector, that is, those that require some measure of dealers to support the high effort equilibrium.[15] The high effort is socially efficient when the associated output compensates for both the non pecuniary costs of exercising this high level of effort, ψ, and the costs of acquiring information, that is

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

The first term of (24) is the output produced by the $1-\widehat{d}_h$ entrepreneurs when they implement the high effort, net of non pecuniary costs. The integral corresponds to the non pecuniary information acquisition costs of the 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 recall, by (8), dominates the allocation where type 2 agents prefer simply to carry their endowment to subsequent dates.

Proposition 7 *Assume it is socially efficient to implement the high effort action, that is,* (24) *holds. Then all equilibria are generically inefficient and, moreover, when the equilibrium features the high action it also features an inefficiently large financial sector.*

It is straightforward to verify that for the parameter values given in section 4.1.1 above the socially efficient origination effort is a_h:

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

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, as in the case in section 4.1.1, the planner wants to support such a level of effort with the minimum measure of dealers $\widehat{d}_h$, for adding "one" additional dealer detracts from productive entreprenurial 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 the financial sector creates a positive externality among dealers via the cream skimming and this leads to a larger financial sector than constrained efficiency would have it.

[15] Recall that by assumption (8)

Assume next that (24) does not hold as it is the case for the parameter values given in section 4.1.2:

$$[\rho(1+a_h(\gamma-1))-\psi]\left(1-\widehat{d}_h\right)-\varphi\widehat{d}_h-\rho(1+a_l(\gamma-1))=-.4408.^{16}$$

In this case, 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 financial sectors and one that indeed supports the constrained social optimum, $(a^*=a_l, d_1^*=0)$.

Finally, the case in section 4.1.3, where there were both low and high effort equilibria, merits some comments as well. First, in this example (24) is not met and thus the high effort is not socially efficient, though it can be supported as a stable equilibrium. There is thus an efficient low effort equilibrium with no financial sector and an inefficient, unstable, one with a strictly positive measure of dealers.[17]

4.2.3 Pareto ranking of multiple equilibria

The previous argument 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 4.1.1, the most efficient equilibrium is the unstable one, d_1^*, which dominates the stable one d_2^*. In the example in section 4.1.2, which dealt 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 origination-effort can be ranked by total ex-ante social surplus in decreasing order of the equilibrium size of the OTC market.*

[16] Here we have used for $\widehat{d}_h$ the value that would obtain were $\overline{d}$ be sufficiently large. When $\overline{d}=.15$, $\widehat{d}_h>\overline{d}$.

[17] There is in principle a fourth case, which we were neither able to rule out nor find an example of, which would consist of a situation where the high effort allocation is socially efficient but where one can support also low effort equilibria, in particular one with $d^*=0$. In this case an equilibrium would be associated with too small a financial sector, rather than one that is too large. Notice though that this case is perfectly consistent with Proposition 7.

5 Competition between dealers

In our model we assumed that an entrepreneur's bargaining power κ is invariant to the number of dealers, d. A natural assumption though is that as the number of dealers increase so does the entrepreneurs bargaining power, that is, that $\kappa(d)$ is an increasing function of d, $\kappa' > 0$. In this brief section we show that the main results of the paper still hold. 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 an inefficiently large dealer sector.

To understand why our main result still holds, perhaps unintuitively, it is useful to return to Proposition 5 and notice that it 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. 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 in turn a simple Corollary to Proposition 5 and thus holds as well when $\kappa' > 0$. This Proposition lies at the heart of the analysis in Section 4. The proposition that does not hold is Proposition 3 as one should no longer expect the utility of a given dealer $\widetilde{d}$ to be monotone in the measure of dealers as now the positive externality is offset, fully or partially, by the effect that competition has on his expected payoff. But this monotonicity is unrelated to our main result. For instance, if a high effort equilibrium exists, stable or unstable, it has to generically feature a measure of dealers that is strictly greater than $\widehat{d}_h$ which is the source of the inefficiency as such an equilibrium will feature a larger measure of dealers than what is needed to support this high effort level. Proposition 7 thus still holds and thus if it is efficient to implement high effort, then all equilibria are generically inefficient and, moreover, when the equilibrium features high effort it must necessarily involve an inefficiently large financial sector.[18]

[18]Note that intuition indicates that the not-very-plausible case of $\kappa' < 0$ also preserves our results. If an

6 Discussion and applications

6.1 OTC markets

Our model offers a simple theory as to why OTC markets arise naturally even in the presence of well functioning exchanges. The reason is that both sides of the market have an incentive to meet outside the exchange, which, once agents are set in their roles, makes both parties better off. Entrepreneurs in distress with good projects may get recognized as such by informed dealers and thus obtain better prices for their assets than otherwise they would in the exchange and dealers can lever up their information cream skimming the good projects. The key here is the bargaining between both parties, the fact that the dealer can bid for the asset in question without being forced to accept any other asset for that bid, as it would happen in a centralized exchange.[19]

The extraordinary profits that dealers[20] obtain in OTC transactions explain the many efforts that broker dealers have made during the discussions leading up to the Dodd-Frank Wall Street Reform and Consumer Protection Act of July 21st, 2010, to prevent OTC contracts from

---

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.

[19]It is for this reason that centralized clearing is *less* of a problem for dealers than execution. For instance Harper, Leising, and Harrington (2009) state that:

> Another debate is over which clearing platforms or exchanges should be used. JPMorgan, Goldman Sachs, Bank of America, Citigroup, Morgan Stanley and other banks will begin sharing profits next year from the credit-default swap clearinghouse ICE US Trust LLC.
>
> While the banks have an interest in supporting that initiative, theyre 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.

Also Leising (2009) states that "[W]hile firms don't want to trade on exchanges, they are willing to have standardized derivatives go through clearinghouses, Pickel said in congressional testimony earlier this year." The issue thus is not standardization but the uniformity of rules and pricing that exchanges impose.

[20]The five biggest derivatives dealers in the US are JPMorgan, Goldman Sachs, Bank of America, Morgan Stanley and Citigroup Inc. Between them they held 95% of the $291 trillion in notional derivatives value of the country's 25 largest bank holding companies at the end of the first quarter of 2009 (see Harper, Leising, and Harrington (2009)).

being forced into exchanges.[21] Cream-skimming, we argue, is a particularly profitable activity for it allows dealers to retain the good risks and leave for the uninformed, but rational, agents the bad ones. For instance derivatives traded in OTC markets, such as credit default swaps, allow dealers to slice cash-flows in a way that makes dealers senior to other claimants and thus leave these with the bulk of the losses when they happen.

Our model also suggests that firms that, for whatever reason, are more likely to draw good projects for any given level of effort and measure of dealers would also prefer to retain OTC markets, rather than force trades into exchanges, as they are more likely to get better terms from dealers than from the organized exchange, where they are more likely to provide the subsidy to lower quality firms in the pooling equilibrium that obtains in these exchanges. By the same token, firms that are more likely to draw low quality projects would rather close OTC markets and force all risks to flow to exchanges, as they benefit from the subsidy. Thus it is not surprising that some "big name" firms do also lobby for keeping OTC markets in their present form.[22] Clearly, and as discussed in Section 5, firms do prefer *more* competition among dealers so that the $\kappa' > 0$ and more of the rents accrue to them rather than to the banks. Conversely, dealers benefit from barring entry to potential competitors and much of the lobbying is directed at preserving these barriers to entry; transparency and quote availability is key in preserving the bargaining power of the dealers and thus their efforts are directed not so much at preventing centralized clearing as maintaining the traditional opacity of execution

[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).

[22] For evidence that not only banks are the ones lobbying for the reservation of OTC markets as they currently stand see Scannell (2009), who states that "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). They add that "[A]t least 42 nonfinancial companies and trade associations are lobbying Congress on derivatives, according to a Wall Street Journal analysis of lobbying disclosure forms filed through April. That's more than triple the 14 nonfinancial companies that lobbied on derivatives in all of 2008 and zero in 2005. The figures include only companies that specifically name derivatives as a lobbying issue." Here the issue is the specific deals the obtain from dealers in the form of tailored contracts, something they fear would be lost were contracts forced into exchanges.

in OTC markets.[23]

6.2 IT and the growth of compensation in the financial industry

Figure 1 shows that the growth in the median compensation of the financial industry was driven by the broker-dealers, which constitute the main entry in Other Finance. These are the main agents that are present in OTC markets and the units inside commercial banks and insurance companies that got richly rewarded during the boom years were precisely those present in these markets, such as AIG's infamous Financial Products group, which made most of its profits trading credit default swaps. But the timing of the growth requires an explanation.

Philippon and Reshef (2008), as well as many others, argue that it is the wave of deregulation that lead to the phenomenal profits and growth of the financial services industry. This, undoubtedly, played a role. But our model offers a different and novel interpretation, which is that improvements in information technology (IT) has effectively decreased the costs of gathering and organizing information in a way that has particularly benefitted agents present in OTC markets, where information has been traditionally dispersed and hard to obtain. In our model this could be captured by an increase in the maximum measure of $\overline{d}$ with finite information gathering costs (see expression (2)). Consider for example the example in section 4.1.1, which showed the case of an economy with two high effort equilibria and focus on the stable one. 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.

[23] As Morgenson (2010) says "Changing the way this market operates is also crucial because major firms like JPMorgan Chase, Goldman Sachs and Morgan Stanley currently run it. These companies don't want to open trading facilities to more participants because they prefer that their customers have limited access to bids and offers; such black-box arrangements generate far more profits to dealers than open and transparent markets."

7 Conclusions

We have presented a model of occupational choice where agents can choose between becoming entrepreneurs and engaging on productive activities or acquiring information and becoming financial intermediaries, 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 relieving moral hazard problems at origination, then we show that the financial markets that arise in equilibrium, and do indeed solve these moral hazard problems, are *always* too large.

Our theory helps explain the rise on 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, the ones that capture these excessively large rents, which is consistent with the observed trends in financial markets. In addition our framework is, we believe, helpful in rationalizing 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 bill in July of 2010 on financial regulatory reform.

REFERENCES

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

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.

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

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, Guillaume Rocheteau and Pierre-Olivier Weill (2009), "Crises and Liquidity in Over the Counter Markets", NBER Working Paper 15414

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 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, 503-530.

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

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

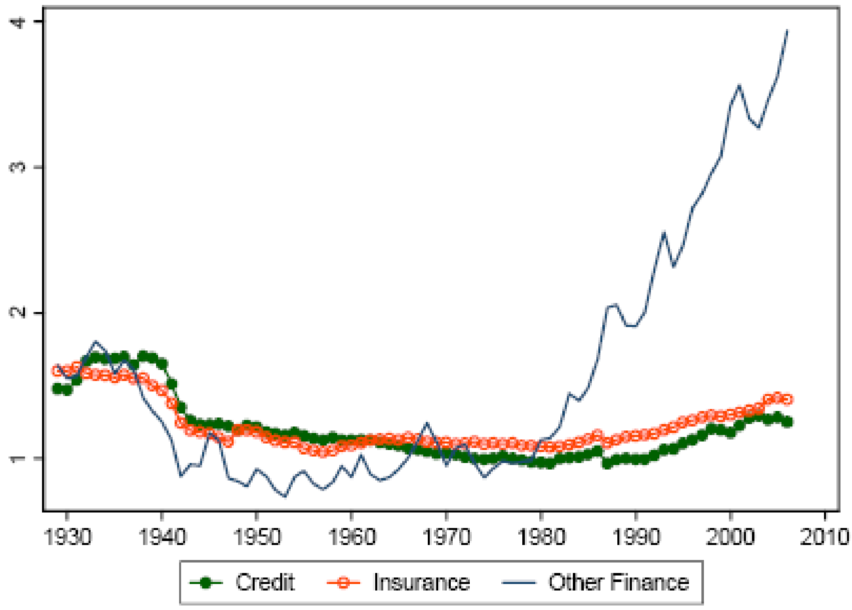

Figure 1: Wages in the financial sector relative to non farm private sector. Figure 2-B in T. Philippon and A. Reshef, "Wages and Human Capital in the US Financial Industry: 1926-2006," NYU working paper, December 2008.

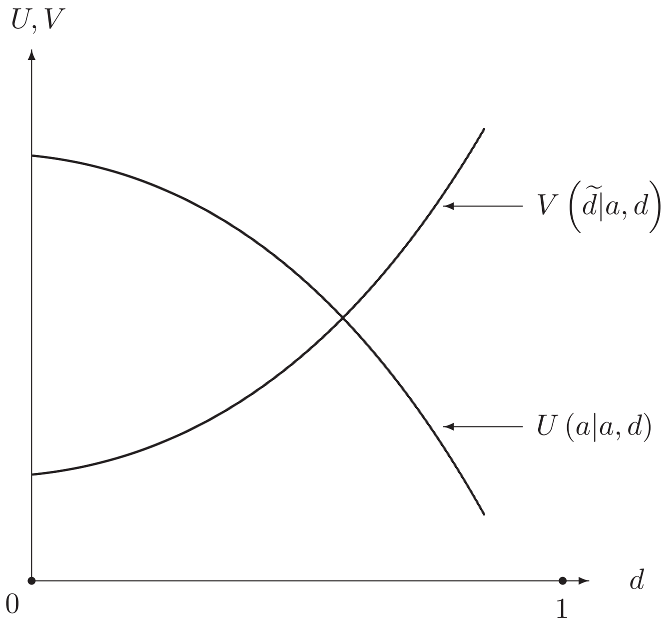

Figure 2: **Payoff functions.** Utility functions of the entrepreneur and a *given* dealer, $\widetilde{d}$, as a function of the measure of dealers, d

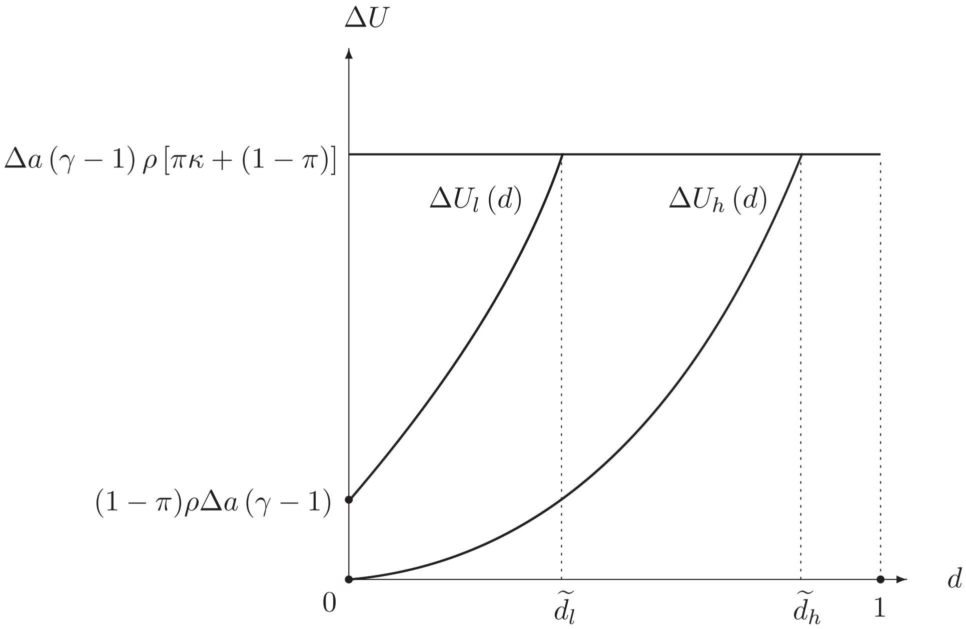

Figure 3: **Incentive compatibility for the low and high effort profile.** The functions $\Delta U_l\,(d)$ and $\Delta U_h\,(d)$ are increasing in d and $\Delta U_l\,(d) > \Delta U_h\,(d)$. $\widetilde{d}_l$ and $\widetilde{d}_h$ are the first measures of dealers for which $m\left(a_l, \widetilde{d}_l\right) = 1$ and $m\left(a_h, \widetilde{d}_h\right) = 1$, respectively.

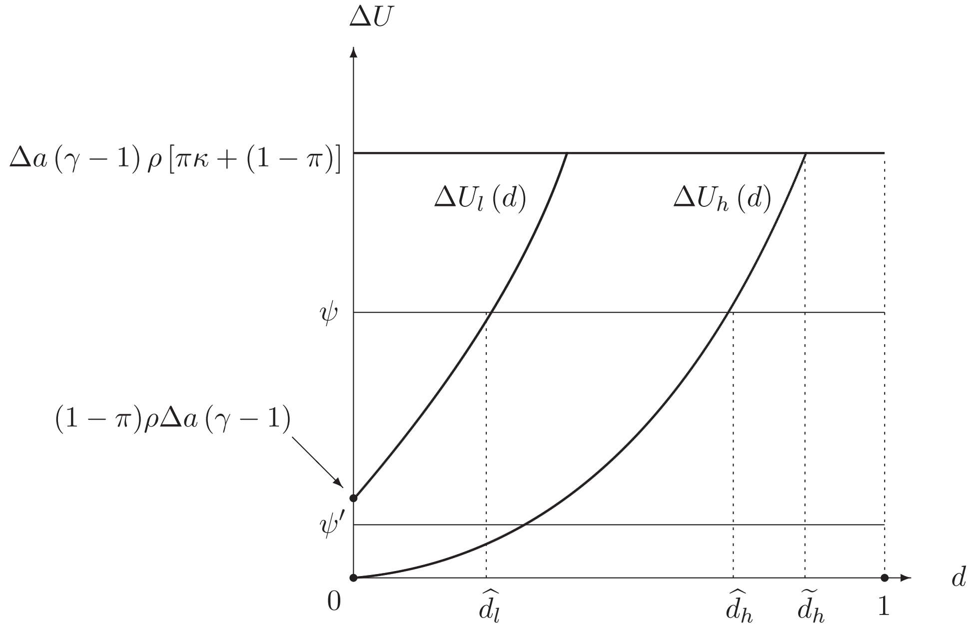

Figure 4: **Incentive compatibility.** Incentive compatibility for the low and high effort profile. When the cost of providing the high effort is given by ψ then a candidate low effort equilibrium is only incentive compatible if and only of $d \in [0, \widehat{d}_l]$ and a candidate high effort equilibrium is only incentive compatible if and only if $d \in [\widehat{d}_h, \widetilde{d}_h]$

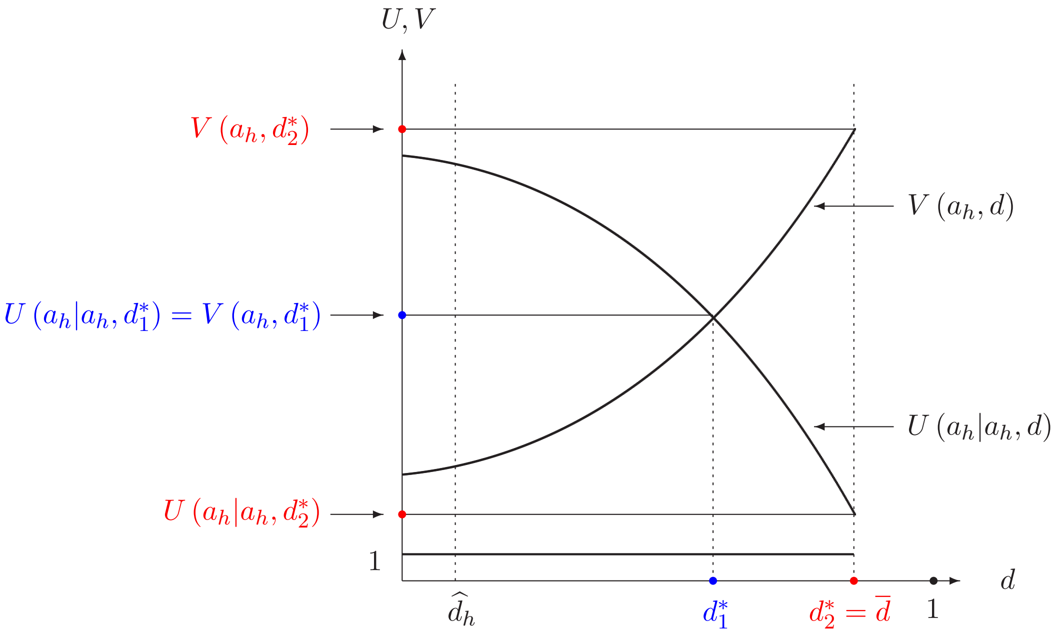

Figure 5: **High effort equilibria.** When the information cost function is as in (22) and the parameter values are given in (23) then $\widehat{d}_h = .0536$ and there are two possible high effort equilibria, one in which $d_1^* = .3101$ where the marginal dealer is indifferent between becoming one or an entrepreneur, and another in which $d_2^* = \overline{d} = .35$ in which the marginal dealer strictly prefers to be one rather than an entrepreneur. Notice also that the allocations meet the participation constraint in that $U(a^*|a^*, d^*) \geq 1$ and $V(d|a^*, d^*) > 1$ for all $d \leq d^*$.

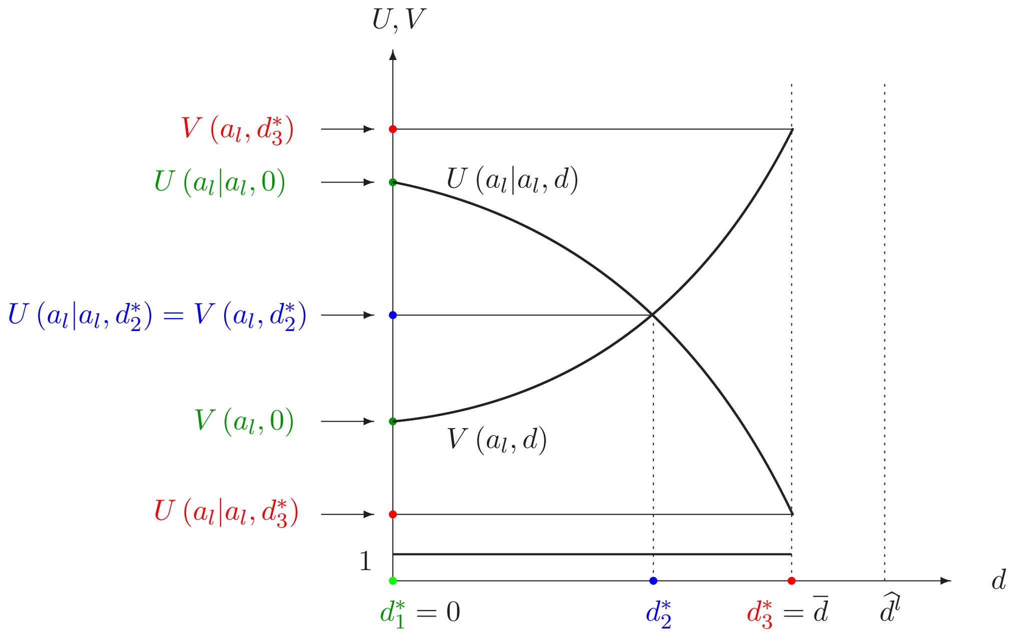

Figure 6: **Low effort equilibria.** When the information cost function is as in (22) and the parameter values are given in section 4.1.2 then $\widehat{d}_l = \overline{d} = .15$ and there are three possible low effort equilibria. First one in which $d_1^* = 0$ where the agent $\widetilde{d} = 0$ prefers to become an entrepreneur rather than a dealer. A second low effort equilibrium is one where $d_2^* = .0781$ and in which the marginal dealer is indifferent between becoming one or an entrepreneur. Finally there is a third low effort equilibrium with $d_3^* = \overline{d} = .15$ in which the marginal dealer strictly prefers to be one rather than an entrepreneur. Notice also that the allocations meet the participation constraint in that $U(a^*|a^*, d^*) \geq 1$ and $V(d|a^*, d^*) > 1$ for all $d \leq d^*$.

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. Such a dealer can only borrow against an asset he has acquired in either the OTC or organized market. Moreover, he can only gain from acquiring and borrowing against an asset if he is able to resell the asset for a profit. But this is not possible in either market at equilibrium prices p and p^d. □

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. □

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

Proof of Proposition 4. Trivial algebra shows that

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

and

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

Finally, tedious computations show that

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

Expressions (25), (26), and (27) are used throughout. □

Proof of Proposition 3. Start with the utility of the dealer, which trivially is such that

$$V_d\left(\widetilde{d}|a,d\right) = \frac{\partial V}{\partial d} = -(1-\pi)(1-\kappa)\,p_d > 0,$$

and

$$V_{dd}\left(\widetilde{d}|a,d\right) = \frac{\partial^2 V}{\partial d^2} = -(1-\pi)(1-\kappa)\,p_{dd} > 0,$$

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

As for the utility of the entrepreneur, start by noticing that tedious algebra yields that

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

Trivial manipulations show that

$$\gamma\rho - p = \gamma\rho - \frac{a(1-m)\gamma + (1-a)}{a(1-m)+(1-a)}\rho = -\left(\frac{a(1-m)+(1-a)}{a}\right)\frac{p_d}{m_d},$$

and hence

$$m_d(\gamma\rho - p) - mp_d = -\left(\frac{a(1-m)+(1-a)}{a}\right)p_d - mp_d = -\frac{p_d}{a},$$

and thus we can write $U_d(a,d)$ as

$$U_d(a,d) = \pi(1-\kappa)p_d < 0.$$

Clearly then,

$$U_{dd}(a,d) = \pi(1-\kappa)p_{dd} < 0,$$

which proves (a). □

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

Proposition 9 *(a) Assume that putative action in equilibrium is $a^* = a_h$ then the utility of the entrepreneur who deviates and chooses instead to exercise action a_l as a function of the measure of dealers d is given by*

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

(b) Assume that putative action in equilibrium is $a^* = a_l$ then the utility of the entrepreneur who deviates and chooses instead to exercise action a_h as a function of the measure of dealers d is given by

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

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

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

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

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

First, notice that

$$U^{\text{sell}}\left(a_l|a_h,d,\text{no-liq.}\right) = a_l m_h(d)\gamma\rho + (1-a_l m_h(d))p_h(d) \tag{30}$$
$$= p_h\left(d\right) + a_l m_h\left(d\right)\left(\gamma\rho - p_h(d)\right),$$

where the functions $p_h\left(d\right)$ and $m_h\left(d\right)$ were given by (10) and (11), respectively, when $a = a_h$. It may be useful to elaborate on (30). The first term 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.

Next notice that

$$U^{\text{no-sell}}\left(a_l|a_h,d,\text{no-liq.}\right) = \rho\left[1 + a_l\left(\gamma-1\right).\right] \tag{31}$$

It is interesting to note that

$$U^{\text{sell}}\left(a_l|a_h,d,\text{liq.}\right) = a_l m_h(d)\left[\kappa\gamma\rho + (1-\kappa)p_h(d)\right] + (1-a_l m_h(d))p_h(d)$$
$$= p_h\left(d\right) + a_l m_h\left(d\right)\kappa\left(\gamma\rho - p_h(d)\right),$$

that is, when the deviant entrepreneur with a good asset is forced to sell due to the occurrence of the liquidity shock he is only able to capture a fraction κ of the surplus $(\gamma\rho - p_h(d))$ when matched with an informed dealer. The difference is precisely that when the deviant agent does not suffer the liquidity shock and receives a bid for the asset, he immediately infers that the asset he holds is a good one and prefers to carry it to maturity.

Notice next that when $d = 0$, $m\left(a_h, d=0\right) = 0$ and

$$U^{\text{sell}}\left(a_l|a_h,d=0,\text{no-liq.}\right) = \rho\left[1 + a_h\left(\gamma-1\right)\right]$$
$$> \rho\left[1 + a_l\left(\gamma-1\right)\right]$$
$$= U^{\text{no-sell}}\left(a_l|a_h,d=0,\text{no-liq.}\right),$$

where we have used the expression $p_h\left(d=0\right)$. Next, define $\widetilde{d}_h$ such that $m\left(a_h,\widetilde{d}_h\right) = 1$. For this measure of dealers, $p_h\left(\widetilde{d}_h\right) = \rho$, and thus

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

Finally notice that $U^{\text{no-sell}}\left(a_l|a_h,d,\text{no-liq.}\right)$ is independent of d and that $U^{\text{sell}}\left(a_l|a_h,d,\text{no-liq.}\right)$ is a decreasing function of d, where the proof of this claim follows exactly the same arguments as the proof in Proposition 3, thus it follows that

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

as we wanted to show. It follows that

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

which after some manipulations yields (28).

(b) Consider now the deviation to a_h when the putative equilibrium features a_l and the measure of dealers is given by d. In this case:

$$\begin{aligned} U^{\text{sell}}\left(a_h|a_l,d,\text{no-liq.}\right) &= p_l(d)+a_h m_l(d)\left(\gamma\rho-p_l(d)\right) \\ U^{\text{no-sell}}\left(a_h|a_l,d=0,\text{no-liq.}\right) &= \rho\left[1+a_h\left(\gamma-1\right)\right]. \end{aligned}$$

We show that in this case, unlike before,

$$U^{\text{sell}}\left(a_h|a_l,d=0,\text{no-liq.}\right)<U^{\text{no-sell}}\left(a_h|a_l,d=0,\text{no-liq.}\right). \tag{33}$$

First notice that at $d=0$, $m_l(d=0)=0$ and thus

$$\begin{aligned} U^{\text{sell}}\left(a_h|a_l,d=0,\text{no-liq.}\right) &= \rho\left[1+a_l\left(\gamma-1\right)\right] \\ &< \rho\left[1+a_h\left(\gamma-1\right)\right] \\ &= U^{\text{no-sell}}\left(a_h|a_l,d=0,\text{no-liq.}\right). \end{aligned}$$

As before define $\widetilde{d}_l$, the measure of dealers for which $m_l\left(\widetilde{d}_l\right)=1$. In this case

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

Finally, notice that

$$U_d^{\text{sell}}\left(a_h|a_l,d,\text{no-liq.}\right)=p_{l,d}\left(1-\frac{a_h}{a_l}\right)>0,$$

as $p_{l,d}<0$, by Proposition 4 and $a_h>a_l$, and this shows that (33) obtains. Thus

$$U\left(a_h|a_l,d\right)=-\psi+\pi\left[p_l(d)+a_h m_l(d)\kappa\left(\gamma\omega\rho-p_l(d)\right)\right]+(1-\pi)U^{\text{no-sell}}\left(a_h|a_l,d,\text{no-liq.}\right). \tag{34}$$

Trivial manipulations of (34) yield (29). □

Proof of Proposition 5. (a) Trivial algebraic manipulations show that

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

as $p_{h,d}<0$ by Proposition 4. Moreover, notice that $\Delta U_h(d=0)=0<\psi$ and that

$$U_h(\widetilde{d}_h)=\Delta a\left(\gamma-1\right)\rho\left[\pi\kappa+(1-\pi)\right],$$

where $\widetilde{d}_h$ is the unique measure of dealers for which $m_h\left(\widetilde{d}_h\right)=1$. Similarly notice that

$$\Delta U_{l,d}=-p_{l,d}\frac{\Delta a}{a_l}\pi\kappa>0,$$

but that now

$$\Delta U_l(d=0)=(1-\pi)\rho\Delta a\left(\gamma-1\right)$$

and that

$$\Delta U_l(\widetilde{d}_l) = \Delta a\,(\gamma - 1)\,\rho\,[\pi\kappa + (1-\pi)] = \Delta U_h(\widetilde{d}_h).$$

(b) As for $\Delta U_h(d) < \Delta U_l(d)$,

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

as

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

by Proposition 4 and this completes the proof. □

Comment. Notice thus that the difference between (28) and (29) is precisely the behavior of the entrepreneur *in the absence of a liquidity shock.* Whereas in the case where the agent deviates to a_l from a_h, he always prefer to sell this is not the case when he deviates to a_h from a_l. Effectively then in the case of the deviation in the putative high effort equilibrium, the utility level under the deviation is higher than the utility of the deviation under the putative low effort equilibrium relative to their respective equilibrium utility levels, which is what s driving the result in Proposition 5 (b). □

Proof of Proposition 6. Given the definition of $\widehat{d}_h$ and $\widehat{d}_l$ in (20) this proposition is an immediate corollary of Proposition 5. □

Proof of Proposition 7. This follows immediately from the observation that all high origination-effort equilibria have a measure of informed dealers $d \in [\hat{d}_h, \overline{d}]$, where $\overline{d}$ was defined in (2) and it was such that $m\left(a, \overline{d}\right) \leq 1$ (see expression (6)). Clearly that there exists an equilibrium for which $d_h^* = \widehat{d}_h$ can only be the case for a set of economies of measure 0, those economies for which the high effort equilibrium measure of dealers, d_h^*, satisfies $\Delta U_h\left(d_h^*\right) = \psi$. Finally notice that for any high effort equilibrium such that $d_h^* > \widehat{d}_h$, there are too many dealers, which can only detract from efficiency, as the measure $d_h^* - \widehat{d}_h > 0$, does not serve any additional incentive purposes and could instead be engaged in entrepreneurial activities. □

Comment. Notice that the arguments in Proposition 7 can be extended to show that even when the high effort is not socially efficient, any low effort equilibrium that features a measure of dealers $d_l^* > 0$ is also socially inefficient. But, if $d_l^* = 0$ can be supported as an equilibrium and (24) is not met, thee is an efficient equilibrium. To put it differently, in our framework the efficient outcome can be supported as an equilibrium only when there is no social role for financial markets.

Proof of Proposition 8. This follows immediately from the last part of the proof of Proposition 7. □