Cream skimming in financial markets[*]

Patrick Bolton

Columbia University and NBER

Tano Santos

Columbia University and NBER

Jose A. Scheinkman

Princeton University and NBER

April 25, 2012

Abstract

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

Cream Skimming in Financial Markets

What does the financial industry add to the real economy? What is the optimal organization of financial markets, and how much talent is required in the financial industry? We revisit these fundamental questions in light of recent events and criticisms of the financial industry. The core issues underlying these questions is whether and how the financial industry extracts excessively *high rents* from the provision of financial services, and whether these rents attract too much talent.[1] Figure 1, from Philippon and Resheff (2008), plots the evolution of US wages (relative to average non-farm wages) for three subsegments of the finance services industry: credit, insurance and 'other finance.' Credit refers to banks, savings and loans and other similar institutions, insurance to life and P & C, and 'other finance' refers to the financial investment industry and investment banks. As the plot shows the bulk of the growth in remuneration in the financial industry took place in 'other finance.'

In this paper we attempt to explain the outsize remuneration in this latter sector by modeling a financial industry that is composed of two sectors: an organized, regulated, standardized, and transparent market where most retail ('*plain vanilla*') transactions take place, and an informal, *opaque* sector, where *informed* transactions take place and '*bespoke*' services are offered to clients. We refer to this latter sector as over-the-counter (OTC) markets[2] and to the transparent, standardized, markets as organized exchanges. A central idea in our analysis is that while OTC markets provide indispensable valuation services to issuers of assets, their opacity also allows informed dealers to extract too high rents. What is more, OTC markets tend to undermine organized exchanges by "*cream-skimming*" the juiciest deals away from them.[3] The informational rents in OTC markets in turn attract too much talent to the financial

---

[1] Goldin and Katz (2008) document that the percentage of male Harvard graduates with positions in Finance 15 years after graduation tripled from the 1970 to the 1990 cohort, largely at the expense of occupations in law and medicine.

[2] Some OTC markets, e.g., markets for foreign exchange, are quite transparent. The important distinction for the present paper is between opaque and transparent markets.

[3] Rothschild and Stiglitz (1976) considered a different form of cream skimming in insurance markets with adverse selection. In that setting, insurers are uninformed about risk types, but offer contracts that induce informed agents to self-select into insurance contracts. For an application of the Rothschild-Stiglitz framework to

industry, which would be more efficiently deployed in the real sector composed, in our model of self-employed entrepreneurs.

The key role of the financial sector in our model is to provide liquidity by allowing entrepreneurs to sell the assets they have originated to investors. These assets vary in quality and a key service provided by the financial industry is valuation of assets for sale. This is where the *talent* employed in the financial industry (specifically, in OTC markets) manifests itself. Importantly, by identifying the most valuable assets and by offering more attractive terms for those assets, informed dealers in the OTC market also serve the role of providing incentives to entrepreneurs to originate good assets. As we argue, however, what matters for the allocative efficiency of talent across the financial and real sectors is what share of the incremental value of good assets dealers get to appropriate. In our model, dealers tend to extract an excessively large informational rent due both to the scarcity of valuation skills (which are costly to acquire) and the opaqueness in OTC markets. What is more, in our model, OTC dealers' rents tend to *increase* as there are more informed dealers, because the greater cream-skimming by dealers worsens the terms entrepreneurs can get for their assets on the organized exchange, and therefore their bargaining power on OTC markets. Our assumption that trading in OTC markets is opaque contrasts with the standard framework first developed by Grossman and Stiglitz (1980). In that class of models, privately produced information leaks out in the process of trading, and as a result *too little* costly information may be produced by 'insiders.' Since many activities in the financial industry can be identified as 'information producing' the Grossman-Stiglitz model seems ideally suited to explain why the financial sector is too small. In contrast, our model helps explain how excessive rent extraction together with excessive entry into the financial industry can be an *equilibrium outcome*.

The coexistence of OTC forwards and futures contracts traded on exchanges provides an interesting illustration for our model. Why don't all future transactions take place on organized futures markets? One reason is as in our model: transactions in forward markets

competition among organized exchanges see Santos and Scheinkman (2001).

2

are primarily between informed dealers and producers who seek to hedge against spot-price movements. By trading in forward markets these producers are typically subject to lower margin calls when the spot price moves away from the forward price. The reason is that informed dealers understand that (as long as they are not over-hedged) producers actually benefit from movements in spot price away from the forward price and therefore do not give rise to higher counterparty risk. As a result, a substantial portion of commodities production is hedged outside exchanges, via forward contracts with banks and trading companies. These contracts give producers less favorable prices, but require smaller margins. After doing due diligence to verify that a producer is not over-hedged, a bank can feel confident that it will actually be better off if spot prices increase. This same bank would most likely also engage in an opposite forward with a counterparty for whom buying forwards would actually lower risk, and only hedge the net amount with futures contracts. Thus, by demanding a uniform mark-to-market margin of all parties, exchanges induce a lower mix of producer-hedgers, and hence a riskier set of buyers and sellers.

Our paper offers a novel hypothesis to explain three related facts about the recent evolution of the US financial services industry, as shown by Philippon and Resheff (2008) and Philippon (2011, 2012). First, the financial services industry accounts for an increasing share of GDP even after financial services exports are excluded - an increase that accelerated starting in the mid 80s. Second, this growth has been accompanied by a substantial increase in IT spending in the financial sector. As Philippon (2012, Figures 5 and 6) shows, other sectors, such as retail, have increased the fraction of spending on IT as well, but in retail there is a *negative* time-series correlation between GDP shares and IT investments. Finally, as already mentioned, there has been a substantial increase in compensation in brokerage and asset management, the segment of finance that is most closely associated with OTC transactions. Our model suggests that developments in IT are partly responsible for these trends. As IT became cheaper, OTC activities which are information intensive became more profitable relative to exchange traded activities. The additional increase in OTC dealers' rents that resulted from the entry of more dealers, provided a reinforcement mechanism for the growth of compensation

in OTC activities and prevented the dissipation of the rents from cheaper IT that was observed in retail. Others have argued that regulatory developments are behind the growth of the OTC sector. Regulatory developments, however, would also be subject to the same reinforcement mechanisms that we argue prevented the dissipation of rents from IT.

The paper is organized as follows: Section 1 outlines the model. Section 2 analyzes entrepreneurs' moral hazard in origination problem and describes some basic attributes of equilibrium outcomes. The analysis of welfare and equilibrium allocation of talent in financial markets is undertaken in section 3. Section 4, in turn, considers the robustness of our main results to the situation where informed dealers compete with each other, or when informed traders are also present on the exchange. The presence of informed traders on the exchange raises the expected price of good assets in the exchange, while lowering the expected price of bad assets. Thus, informed traders on the exchange also provide incentives to originate good assets and they dampen the effects of cream-skimming by OTC dealers. We show that in this more general and realistic situation OTC markets are more likely to be excessively large in equilibrium. Section 5 concludes. All proofs are in the Appendix.

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

In contrast to this literature, our model emphasizes the fifth function in Levine's list: secondary market trading and liquidity provision. In addition, where the finance and growth

4

literature only distinguishes between bank-based and market-based systems (e.g. Allen and Gale, 2000), a key distinction in our model is between markets in which trading occurs on a bilateral basis at prices and conditions that are not observable by other participants, and organized exchanges with multilateral trading at prices observed by all.[4]

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

---

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

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

of these markets and they do not address issues of cream-skimming or occupational choice.

1 The model

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

1.1 Agents

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

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

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

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

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

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

In all other respects, type 2 agents are identical: They face the same *i.i.d.* liquidity shocks: they value consumption only in period 1 with probability $0 < \pi < 1$ and only in period 2 with probability $(1 - \pi)$.[6] Their preferences are represented by the utility function

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

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

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

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

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

If type 2 agent d chooses to work in the financial sector as a dealer, he saves his unit endowment to period 1, but incurs a utility cost $\varphi(d)$ to build up human capital in period 0. This human capital gives agent d the skills to value assets originated by entrepreneurs and that are up for sale in period 1. Specifically, we assume that a dealer is able to perfectly ascertain

---

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

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

1.2 Financial Markets

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

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

Consider first the OTC market. This market is composed of a measure $d(1 - \pi)$ of *patient* dealers ready to buy assets from $(1 - d)\pi$ *impatient* entrepreneurs. Each dealer is able to trade a total output of at most $1 + \rho$, his endowment plus a maximum collateralized loan from rentiers of ρ, in exchange for claims on entrepreneurs' output in period 2. Impatient entrepreneurs turn to dealers for their information: they are the only agents that are able to tell whether the entrepreneur's asset is worth $\gamma\rho$ or just ρ. Just as in Grossman and Stiglitz (1980), dealers' information must be in scarce supply in equilibrium, as dealers must be compensated

---

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

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

for their cost $\varphi(d)$ of acquiring their valuation skills. As will become clear below, this means not only that dealers only purchase high quality assets worth $\gamma\rho$ in equilibrium, but also that not all entrepreneurs with high quality assets will be able to sell to a dealer.

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

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

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

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

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

---

[9]The assumption is formally made below.

[10]For a similar approach to modeling negotiations in OTC markets between dealers and clients see Lagos,

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

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

or

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

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

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

Consider next the organized exchange. We show that in equilibrium all assets of impa-

---

Rocheteau, and Weill (2010).

[11] In Section 4 we show that our results are robust to assuming that the bargaining power of dealers decreases with the number of dealers.

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

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

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

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

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

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

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

1.3   Discussion and parameter restrictions

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

In our model, entrepreneurs have an added incentive to choose high effort because 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. If it were not for these incentive effects, informed dealers would enrich themselves thanks to their cream-skimming activities, but would not create any social surplus.

11

We introduced heterogeneity among type 2 agents only in the form of different utility costs to become a dealer. We could also have introduced heterogeneity in the costs of becoming an entrepreneur. We would then simply order type 2 agents in their increasing *comparative advantage* of becoming dealers and proceed with the analysis as in our current model.

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

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

$$\gamma\rho < 1 + \rho. \tag{6}$$

In addition, and in order to simplify the presentation in what follows, we restrict ourselves to situations where even in the absence of a dealer sector, $d = 0$, type 2 agents would prefer to become entrepreneurs and exercise the low effort rather than simply carry their en-

dowments forward. We show in the appendix that to obtain this it is enough to assume that

$$\rho\left[1+a_l\left(\gamma-1\right)\right]>1. \tag{7}$$

1.4 Definition of equilibrium

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

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

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

2 Equilibrium payoffs and the moral hazard problem

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

2.1 Equilibrium borrowing and asset trading in period 1

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

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

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

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

14

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

2.2 Equilibrium payoffs in period 0

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

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

$$U(a^*|a^*, d) = -\psi(a^*) \quad + \quad \pi\left[a^* m(a^*, d) p^d(a^*, d) + (1 - a^* m(a^*, d)) p(a^*, d)\right] \quad (8)$$
$$+ \quad (1 - \pi)\rho\left[1 + a^*(\gamma - 1)\right],$$

In equation (8), $-\psi(a^*)$, 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 events fails to occur, an event with probability $1 - a^* m(a^*, d)$, then the agent needs to sell his project in the 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 formulas for p^d, p and m are given in Section 1.2.

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

$$V\left(\widetilde{d}|a^*, d\right) = -\varphi\left(\widetilde{d}\right) + 1 + (1 - \pi)(1 - \kappa)(\rho\gamma - p(a^*, d)). \quad (9)$$

15

The first term in (9), $-\varphi(\widetilde{d})$, 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, $p(a^*, d)$.

We next characterize how utilities depend on the measure of dealers d.

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

It is useful to consider first the following result, which is immediate.

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

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

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

16

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

Another implication of our model is that fixing the number of dealers d, dealers prefer an equilibrium with low effort, because for a given d, the proportion of good projects and consequently the price in the exchange is lower under low than under high effort. Thus if dealers could induce more bad asset origination, they would do so.

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

2.3   Entrepreneur moral hazard

The action a^* prescribed in equilibrium must be *incentive compatible* that is,

$$U(a^*|a^*, d^*) \geq U(a|a^*, d^*) \qquad for \qquad a \neq a^*. \tag{10}$$

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

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

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:

$$U_h(d) = -\psi + \pi \left[a_h m_h(d) p_h^d(d) + (1 - a_h m_h(d)) p_h(d) \right] + (1 - \pi)\rho \left[1 + a_h (\gamma - 1) \right],$$
(11)

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

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

$$U_{hl}(d) = p_h(d) + a_l m_h(d) (\gamma \rho - p_h(d)) (\pi \kappa + (1 - \pi)).$$
(12)

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

$$\Delta U_h(d) \;=\; \psi + U_h(d) - U_{hl}(d)$$
(13)

18

$$= \pi \Delta a m_h \left(d \right) \kappa \left(\gamma \rho - p_h \left(d \right) \right)$$
$$+ \quad \left(1 - \pi \right) \left[\rho \left(1 + a_h \left(\gamma - 1 \right) \right) - \left(p_h \left(d \right) + a_l m_h \left(d \right) \left(\gamma \rho - p_h \left(d \right) \right) \right) \right].$$

Incentive compatibility requires that

$$\Delta U_h \left(d \right) \geq \psi. \tag{14}$$

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

$$U_l \left(d \right) = \pi \left[a_l m_l \left(d \right) p_l^d \left(d \right) + \left(1 - a_h m_l \right) p_l \left(d \right) \right] + \left(1 - \pi \right) \rho \left[1 + a_l \left(\gamma - 1 \right) \right] \tag{15}$$

where $p_l^d \left(d \right)$, $p_l \left(d \right)$, and $m_l \left(d \right)$ are defined as before, with the obvious changes in notation.

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

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

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 $\Delta U_l \left(d \right)$ denotes the difference in expected *monetary payoffs* (not accounting for effort costs ψ) between the utility under the deviation and the utility that obtains if the agents sticks to the candidate equilibrium action a_l:

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

Incentive compatibility requires that

$$\Delta U_l \left(d \right) \leq \psi. \tag{16}$$

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

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

The functions $\Delta U_h(d)$ and $\Delta U_l(d)$ are shown in Figure 2. There are two reasons why these functions are increasing in d. First, a greater mass of dealers increases the likelihood $m(a^*, d)$ that an entrepreneur with a good asset is matched with an informed dealer. Second, 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 sells bad (and some good) assets.

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

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

$$\widehat{d}_h = \inf\{d \leq \bar{d} : \Delta U_h(d) \geq \psi\} \qquad and \qquad \widehat{d}_l = \sup\{d \leq \bar{d} : \Delta U_l(d) \leq \psi\} \qquad (17)$$

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

Proposition 6 is key in establishing the main results of the paper. In Figure 2 we consider two possible costs of exercising the high effort, ψ and ψ'. If the high effort is socially optimal, and we provide a condition below under which this is the case, then the existence of an OTC market of at least size $\widehat{d}_h$ is necessary to support it. Even when the cost of exercising the high effort is arbitrarily small this effort level is never incentive compatible when d is close to 0. The reason is that, under the candidate high effort equilibrium, the price of the asset in the uninformed exchange is very high when d is close to 0. There is a large measure of

20

entrepreneurs, $1 - d$, all exercising the high effort and there is little cream skimming and hence the quality of the assets in the exchange is high. Thus the price in the uninformed exchange is close to $[1 + a_h (\gamma - 1)] \rho$, the price the asset commands in the absence of any cream skimming. An agent deviating to low effort, if not receiving an offer from an informed dealer, will be able to sell the asset at $t = 1$, independently of whether he suffers a liquidity shock, for a price higher than his uninformed private valuation.[14] Also because there are few informed dealers the entrepreneurs have little hopes of being matched to them at date 1 and thus of capturing some of the surplus $\gamma \rho - p(d)$; thus, given that his high effort provision is likely to go unrewarded in case of distress, the agent prefers simply to save on effort costs and free ride on the large pool of entrepreneurs exercising the high effort.

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

3   Allocation of talent and welfare

3.1   The equilibrium size of the financial and real sectors

We now turn to a central question of our analysis: What is the optimal allocation of talent to the financial sector? Is there too much information acquisition in financial markets? In our model, these questions boil down to determining whether the equilibrium measure of dealers d^* is too

---

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

large. As we saw in Proposition 6, a low (high) effort equilibrium can only be supported when $d \leq \widehat{d}_l$ (*resp.* $d \geq \widehat{d}_h$). Low effort equilibria thus are associated with relatively small financial sectors when compared with high effort equilibria.

It is relatively simple to construct examples for which there is no symmetric equilibrium and for which there are multiple ones. Rather than provide a full characterization of the many possible cases, we provide examples of three possible cases: One in which there are only high effort equilibria, one in which there are only low effort equilibria and one in which low and high effort equilibria coexist. Recall also that for a particular (a^*, d^*) to be an equilibrium a^* must be incentive compatible and, given (18), d^* has to be such that

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

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

$$
\varphi\left(d\right) = \overline{\varphi} \quad for \quad d < \overline{d} \qquad and \qquad \varphi\left(d\right) = +\infty \quad for \quad d \geq \overline{d}, \tag{18}
$$

Under (18) all dealers have identical costs and thus when plotting the expected payoff function of one of them we also plot that of the *marginal* dealer, who determines the size of the OTC market. We may thus define $V\left(a, d\right) := V(\widetilde{d}|a, d)$, for any $\widetilde{d} \leq d < \overline{d}$.

High effort equilibria: Consider the following parameter values

$$
a_h = .75 \quad a_l = .55 \quad \gamma = 1.5 \quad \rho = .8 \quad \kappa = .25 \quad \pi = .5. \tag{19}
$$

We also choose

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

In this case, $m < 1$ for $d = \overline{d}$. There is no low effort allocation that is incentive compatible in this example since $(1 - \pi)\rho\Delta a\left(\gamma - 1\right) > \psi$. High effort is incentive compatible if $d \geq \widehat{d}_h = .0536$. There are two high effort equilibria and they are shown in Figure 3. There is an unstable equilibrium with $d_1^* = .3106$ in which all agents $d \leq \overline{d}$ are indifferent between

becoming entrepreneurs or dealers. There is also a stable equilibrium with $d_2^* = \bar{d} = .35$, in which dealers are *strictly* better off than entrepreneurs. Notice that all agents who can become dealers at a finite cost are dealers in this equilibrium.

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

Low effort equilibria: Suppose (19) holds but

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

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

Coexistence of high and low effort equilibria: Suppose now that

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

while the rest of the parameters are as in (19). There are three equilibria, two that feature low effort and one stable high effort equilibrium.

Then $\widehat{d}_l = .0545$ and there are two low effort equilibria. A stable one with $d_1^* = 0$, since $U(a_l \mid a_l, 0) > V(0, a_l, 0)$, and an unstable equilibrium where type 2 agents with $d \leq \bar{d}$ are indifferent between becoming dealers or entrepreneurs and $d_2^* = .05$. In this example $\widehat{d}_h = .4020$. The allocation $(a_h, d_3^* = .41)$ is a stable high effort equilibrium.

3.2 Welfare: Are OTC markets too large?

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

23

allow the planner to dictate the occupation of type 2 agents but we do not let the planner make any decisions based on the information obtained by dealers. Given a vector of parameters $\mathcal{A} = (a_h, a_l, \gamma, \rho, \kappa, \pi, \psi)$ and the cost function $\varphi(d)$,[15] the planner chooses d knowing that trade will occur in time 1 in the OTC market with d dealers and in the organized exchange at equilibrium prices. The planner's problem in period 0 is then to pick the measure d of type 2 agents that maximizes ex-ante social surplus. Since type 1 agents get no surplus in equilibrium the planner only has to weight the utility of type 2 agents, and we assume that all type 2 agents receive equal weight. If the planner wishes to implement low effort, the optimal choice is obviously $d = 0$ which yields a total surplus that equals $\rho(1 + a_l(\gamma - 1))$. If the planner chooses to implement high effort, she must choose a $d \geq \widehat{d}_h$ and this yields surplus:

$$[\rho\left(1 + a_h\left(\gamma - 1\right)\right) - \psi]\left(1 - d\right) - \int_0^d \varphi\left(u\right) du, \tag{20}$$

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

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

$$[\rho\left(1 + a_h\left(\gamma - 1\right)\right) - \psi]\left(1 - \widehat{d}_h\right) - \int_0^{\widehat{d}_h} \varphi\left(u\right) du \geq \rho\left(1 + a_l\left(\gamma - 1\right)\right). \tag{21}$$

The first term of (21) is the output produced by the $1 - \widehat{d}_h$ entrepreneurs when they implement the high effort, net of costs. The integral corresponds to the information acquisition costs of type 2 agents who become dealers. The high effort is socially efficient if this term is more than what society would obtain if all type 2 agents become entrepreneurs and perform the low effort, which by (7) dominates the allocation where type 2 agents prefer to simply carry their endowment to subsequent dates. Of course, if a high effort equilibrium exists, it is unlikely that $d_h^* = \widehat{d}_h$. The next proposition states this fact more precisely. To treat perturbations to the cost function φ we will consider a family $\varphi^\beta(d) = \varphi(d) + \beta$. Given $\mathcal{A}$ and φ we will show that we can always find parameter values $(\beta, \mathcal{B})$ "close to" $(0, \mathcal{A})$ for which all equilibria for

---

[15]Our assumptions imply that $\mathcal{A}$ is restricted to an open set in $\Re^{7+}$.

parameters sufficiently close to the new values $(\beta, \mathcal{B})$ are constrained inefficient. Thus the set of "bad" parameters is open and dense.

Proposition 7 *Suppose that φ is a smooth function in $(0, \bar{d})$ and that it is socially efficient to implement the high effort action; that is, inequality (21) holds. Then given any set of parameters $\mathcal{A}$ and any $\epsilon > 0$, there exists a $0 \leq \beta < \epsilon$ and vector $\mathcal{B}$ with $|\mathcal{B} - \mathcal{A}| < \epsilon$ and such that for all parameter values sufficiently close to $(\beta, \mathcal{B})$ all equilibria are inefficient and any high effort equilibrium features too many dealers in OTC markets.*

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

It is easy to check that in the first example in Section 3.1 the socially efficient origination effort is a_h and both equilibria feature an excessively large financial sector ($d > \widehat{d_h}$). Conditional on a_h being efficient, the planner wants to support this level of effort with the minimum measure of dealers $\widehat{d_h}$, for adding "one" additional dealer detracts from productive entrepreneurial activities and does not improve incentives. However, this level is not an equilibrium - entry into OTC markets creates a positive externality among dealers via the cream skimming and this leads to a larger OTC market than constrained efficiency would have it.

In the second example in Section 3.1, the constrained efficient allocation calls for a_l and $d = 0$. Notice that in that case there were three equilibria, two of which feature excessively large OTC markets and one that supports the constrained social optimum, $(a^* = a_l, d_1^* = 0)$.

In the last example Section 3.1, (21) is not met and thus high effort is not socially efficient, though it can be supported as a stable equilibrium. There is also an efficient low effort equilibrium and an inefficient one with a strictly positive measure of dealers.

The argument above highlights that, conditional on a particular level of effort, equilibria can be Pareto ranked in decreasing order of the measure of dealers. Thus in the first example

in Section 3.1, the most efficient equilibrium is the unstable one, d_1^*, which dominates the stable one d_2^*. In the second example in Section 3.1, $d_1^* \succ d_2^* \succ d_3^*$. In fact,

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

4 Extensions

4.1 Competition between Dealers

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

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

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

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

Proposition 6, which describes the set of possible measures of dealers in the low and high effort equilibria, is a Corollary to Proposition 5, and thus holds as well when $\kappa' \geq 0$. This Proposition lies at the heart of the analysis in Section 3. Proposition 3 however no longer holds; the positive externality may be offset by the effect of greater competition on κ. But the monotonicity of each dealer's utility with respect to d is unrelated to our main result. For

instance, if a high effort equilibrium exists, it generically features a measure of dealers that is *strictly* greater than $\widehat{d}_h$, which is *inefficient*. Proposition 7 thus still holds.[16]

4.2 Information in the public exchange

In this section we explore the consequences of allowing for informed trading both in the OTC market and on the organized exchange. To this end we extend the model by introducing two choices for informed intermediaries: To become an OTC dealer incurring a personal cost $\varphi(d)$ or an *informed exchange trader* who incurs a smaller cost $\lambda\varphi(d)$, with $\lambda < 1$. Each type of intermediary can determine perfectly the value of the asset and the key difference between the two types is the context in which they trade. OTC dealers trade in an opaque market and their offers are not publicly disclosed. In contrast, dealers trading in the exchange have to disclose their quotes. As a result, their private information may be (partially) inferred by uninformed investors, who can revise their bids in light of this information and therefore compete with informed traders. The assumption that $\lambda < 1$ is to reflect a higher fixed costs involved in over-the-counter trading and to allow for the coexistence of dealers and informed traders. To highlight the trade-off between dealers and informed traders and simplify other aspects of the model we will assume that the fraction of entrepreneurs is fixed. In addition we will also assume that only entrepreneurs that receive a liquidity shock put their projects for sale.[17]

Over-the-counter dealers trade exactly as in our base model. We consider the following trading protocol in the organized exchange. First, all assets are put up for sale simultaneously at some price p_u. Any buyer willing to bid more than p_u can make a targeted bid for a specific asset; all these bids are public information. Then, any other buyer can submit counterbids on these targeted assets using what they inferred from the first round of bidding, after which all assets are sold to the highest bidder.

---

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

[17]In the main model only entrepreneurs that receive a liquidity shock put their projects for sale in equilibrium.

In the absence of any additional signals the most an uninformed buyer is willing to bid is p_u. Thus, if an informed trader bids more than p_u to secure the purchase of a good asset that he has identified, his information would leak out to all buyers and he will face competition from uninformed buyers. If this information leaks out perfectly, then uninformed buyers are willing to bid up to $\gamma \rho$ (the value of the good asset that has been identified), thus completely bidding away the informed trader's information rent. If that is the case, no costly information about asset values will be produced; there will be no 'price discovery' in equilibrium in the organized exchange.

To avoid this outcome we follow the literature spawned by Grossman and Stiglitz (1980) and introduce a form of noise traders adapted to our model, which we refer to as uninformed *'noise buyers'*. Each noise buyer makes a bid $\phi \rho$, with $\gamma > \phi > 1$, fon an asset for which she has an especially high private consumption value. We assume informed traders move first by bidding for valuable assets and that noise buyers move second by bidding for an asset they particularly favor. Uninformed traders only see that an asset received a bid and if, in equilibrium, informed and noise buyers bid the same, the uninformed cannot tell whether the bid is from an informed or a noise buyer. By bidding precisely $\phi \rho$ informed traders can hide behind noise traders and partially protect their information. To simplify our expressions we assume that no asset receives multiple bids.

Let d be the fraction of type II agents that choose to become dealers and i the fraction that chooses to become informed traders. The fractions d and i will be determined in equilibrium, but the total $f = d + i$ is given, so that the total number of projects originated by entrepreneurs is $1 - f$. Let μ denote the measure of noise buyers.

In our candidate equilibrium informed traders and noise buyers make targeted bids $\phi \rho$ for assets. To ensure that uninformed buyers cannot gain by making bids to targeted projects we must assume that the mass μ of noise buyers is large enough that the expected value of targeted projects is less than $\phi \rho$.

As before, entrepreneurs with liquidity needs first pursue the free option to sell their asset in the OTC market. If they have a good project they will receive a (weakly) more attractive

28

bid with probability m_d, and if they don't obtain a bid they can always put their asset up for sale on the exchange. Thus, after OTC dealers have cream-skimmed a mass of good assets $d(1-\pi)$ there are respectively

$$q_g = a\pi[1-f] - d(1-\pi) \qquad and \qquad q_b = \pi(1-a)[1-f], \tag{22}$$

good and bad projects for sale on the exchange. For simplicity, we shall assume that there are enough good projects for sale to meet the demands of both dealers and informed traders:

$$a\pi(1-f) - d(1-\pi) > i(1-\pi)$$

or

$$a\pi(1-f) - f(1-\pi) > 0. \tag{23}$$

After informed traders have bid, the remaining pool of assets has proportions

$$\omega_g = \frac{\pi a(1-f) - (1-\pi)f}{\pi(1-f) - (1-\pi)f} \qquad and \qquad \omega_b = \frac{\pi(1-a)(1-f)}{\pi(1-f) - (1-\pi)f}. \tag{24}$$

of good and bad projects. Note that noise buyers buy random projects and therefore do not affect the proportion of good and bad assets for sale. Thus in equilibrium, the price paid by (risk-neutral) uniformed buyers is

$$p_u = \omega_g \gamma \rho + \omega_b \rho, \tag{25}$$

which only depends on the size of the financial sector f, and not in the relative number of dealers and informed traders.

To insure that uninformed investors do not want to out-bid targeted bids we assume that the mass μ of noise buyers is large enough so that:

$$\left(\frac{\mu}{\mu+i}\right)p_u + \left(\frac{i}{\mu+i}\right)\gamma\rho < \phi\rho. \tag{26}$$

The left hand side of condition (26) is the expected value of a targeted asset for an uninformed investor, which is lower than the cost $\phi\rho$ under condition (26).

A good asset for sale on the organized exchange gets a bid $\phi\rho$ from an informed trader with probability

$$m_i = \frac{(1-\pi)\,i}{a\pi\,(1-f) - (1-\pi)\,d} \quad . \tag{27}$$

If a good asset for sale in the exchange does not get a bid from an informed trader, it would get a bid from a noise buyer with probability

$$m_n = \frac{(1-\pi)\,\mu}{\pi\,(1-f) - (1-\pi)\,f} \quad . \tag{28}$$

Thus, the expected value of a selling a good asset on the exchange is given by

$$p = m_i\phi\rho + (1-m_i)\{m_n\phi\rho + (1-m_n)\,p_u\}, \tag{29}$$

and the price that a dealer pays an entrepreneur with a good asset on the OTC market is:

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

If we hold the size of the financial sector f constant, but increase the relative number of OTC dealers then m_i decreases and m_n stays constant. Therefore p, the expected value of selling a good asset on the exchange, also decreases. In other words, a shift of informed buying to the OTC market away from the exchange worsens the terms (whether it is p or p_d) at which entrepreneurs can hope to sell good assets, and therefore increases the informational rents of dealers on the OTC market. This observation is formalized in the next proposition.

Proposition 9 *The terms of trade p and therefore also p_d are a decreasing function of the number of dealers d.*

If $p_d < \phi\rho$ then entrepreneurs with good projects are worse off when there is a switch of informed trading from the exchange to the OTC market. We will show that in any equilibrium with a strictly positive number of dealers d and informed traders i, $p_d < \phi\rho$. As a result, in any such equilibrium there are too many dealers from the perspective of social efficiency.

The shift in composition of informed trading on and off the exchange isolates the cream-skimming pecuniary externality from the rise in informed trading on the OTC market more

accurately than in our benchmark model. Here the only margin is whether more trading occurs on or off the exchange and everything else, whether it is the total volume of informed trading or the mass of assets that are originated, is held constant. As we show, a greater shift towards OTC markets is unambiguously a negative externality resulting in a welfare loss.

To see this, consider the situation where it is socially optimal for entrepreneurs to implement the high effort, and where accordingly the number of dealers and informed traders (d, i) must be large enough; more formally, (d, i) must then belong to a certain (closed) subset $\mathcal{S} \subset \{(d, i) \in \mathbb{R}_+^2 : d + i = f\}$. The most efficient way of implementing the high effort is then to choose $\hat{d} = \inf\{d : \text{there exists } i \text{ with } (d, i) \in \mathcal{S}\}$ and make each $d \leq \hat{d}$ a dealer and each $\hat{d} < d \leq f - \hat{d}$ an informed trader. This latter observation follows from the relative costs of producing informed traders and dealers.

In an equilibrium with a strictly positive number of dealers and informed traders,[18] the marginal agent d^* must be indifferent between becoming a dealer or an informed trader:[19]

$$-\varphi\left(d^*\right) + 1 + (1 - \pi)\left(\gamma\rho - p_d(d^*)\right) = -\lambda\varphi\left(d^*\right) + 1 + (1 - \pi)\left(\gamma - \phi\right)\rho$$

or,

$$(1 - \pi)\phi\rho = (1 - \pi)p_d\left(d^*\right) + (1 - \lambda)\varphi\left(d^*\right). \tag{30}$$

Now unless $d^* = 0$, equation (30) can only hold if $p_d\left(d^*\right) < \phi\rho$. It follows that if an equilibrium $d^* > 0$ implements the high effort, so that $(d^*, f - d^*) \in \mathcal{S}$, this equilibrium must be inefficient: a small decrease in the number of dealers, holding f constant, makes entrepreneurs with good projects better off, while the payoff of entrepreneurs with bad projects remains unchanged, so that entrepreneurs have a strictly higher incentive to provide high effort. Hence $d^* > \hat{d}$. We summarize this argument in the following proposition.

Proposition 10 *When informed trading takes place on or off the exchange, then any high-effort equilibrium is such that there are too many dealers trading on the OTC market.*

---

[18]In the appendix we exhibit a robust example of such an equilibrium.

[19]The solution is not necessarily unique since the right hand side of (30) is not monotone in d.

It is worth emphasizing that this result is stronger than Proposition 8 established for the benchmark model. The logic behind this result is particularly simple and compelling: Any shift in informed trading away from the exchange and onto the OTC market results in a worsening of the terms of trade for entrepreneurs with good assets and therefore undermines incentives towards origination of good assets. This reduction in origination incentives must then be compensated with a larger number of informed dealers to maintain incentives towards high effort, which means an efficiency loss.

4.3   Growth of OTC Markets and the Role of Information Technology

As we highlight in Bolton, Santos and Scheinkman (2012), the decade that preceded the financial crisis was a period of abnormal growth in the size of OTC derivatives, swaps, commodities, and forward markets. Similarly, Philippon and Resheff (2008) have shown that the abnormal growth in median compensation in the financial industry since the early 1980s is driven in large part by the compensation of broker-dealers, which constitute the main entry in their 'other finance' category (see Figure 1 below). Broker-dealers, of course, are the main players in OTC markets along with units inside commercial banks and insurance companies, such as AIG's infamous Financial Products group, which have been richly rewarded during the boom years prior to the crisis.

What explains the growth in this sector and its timing? Our analysis is cast in a static model, which cannot lend itself to a dynamic explanation of this phenomenon. Still, a simple comparative static exercise in our model can shed light on a combination of factors that surely has facilitated this transformation in the financial sector - conceptual innovations in the valuation of financial derivatives that along with technological advances in information technology (IT) have decreased the costs of processing financial transactions, bookkeeping, and product innovation in decentralized markets. As MacKenzie (2006) suggests, it is not just the development of the Black-Scholes option pricing formula which has made it easier to value and trade financial derivatives, but also the inception of personal computers, as well as electronic

trading and bookkeeping. The reliance on increasingly powerful IT tools has enabled dealers in OTC markets to offer more and more sophisticated and customized financial products and to process a huge volume of transactions. In our very stylized model, this IT revolution can be simulated by a simple comparative statics exercise: an increase in the number of agents that have a low cost of becoming a dealer from $\bar{d}$ to $\bar{d} + \varepsilon$. This increase leads to a new stable high effort equilibrium in the example in section 3.1.1, with a larger OTC market (by an amount ε), higher compensation for dealers, and lower ex-ante profits for entrepreneurs.[20]

Two examples provide a simple illustration of the role IT technology has played in financial innovation and customization. The first is commodities forward contracts, which have been increasingly geographically customized in recent years thanks to satellite imaging technology and IT applications such as Google Earth. Due to their more accurate geographic footprint, these contracts offer more valuable insurance, which in turn enables dealers in these contracts to extract higher profits. The second is energy derivatives such as those offered by Enron Capital and Trade Resources (ECT) a subsidiary of Enron, which set up a "gas bank"– essentially a financial intermediary between buyers and sellers of natural gas–offering both price stability and local gas-supply and demand assurance. As Tufano (1996, pp 139) puts it "ECT's risk managers have clear instructions to develop a hedging strategy that minimizes net gas exposures, and the company has invested millions of dollars in hardware, software, and bundreds of highly trained personnel to eliminate mismatches and ensure that fluctuations in gas priees do not jeopardize the company's existence."

It is interesting to recall that before its eventual collapse, Enron, and in particular ECT, received numerous awards for these innovations.

---

[20]Deregulation, which some commentators (*e.g.* Philippon and Resheff, 2008) suggest was responsible for the phenomenal growth of the financial services industry in the past quarter century, would have a similar effect: the decrease in costs in OTC activities would generate a larger OTC market, higher compensation for dealers, and lower ex-ante profits for entrepreneurs.

5 Conclusions

We have presented a model where individual agents can either work in the real sector and engage in productive activities, or in the financial sector and provide liquidity as well as valuation services. We have asked whether in such an occupational choice model the equilibrium size of the financial sector is efficient. We have identified a novel externality, cream skimming in OTC-like markets, that tends to generate an inefficiently large OTC sector, in which dealers are overly compensated for their valuation and liquidity-provision services.

Our theory helps explain the simultaneous growth in the size of the financial services industry and the compensation of dealers in the most opaque parts of the financial sector. OTC markets emerge even in the presence of well functioning exchanges. The reason is that both entrepreneurs and informed dealers have an incentive to meet outside the exchange: Entrepreneurs with good projects may get better offers from informed dealers than are available on the exchange, and dealers can use their information to cream-skim good projects. Our model thus offers a novel theory of endogenous segmentation of financial markets, where "smart-money" investors deal primarily in opaque OTC markets to protect their information, and uninformed investors trade on organized exchanges. This is in contrast with models in the vein of Grossman and Stiglitz (1980), where instead smart-money investors are assumed to be trading on the exchange, and where as a consequence too little (costly) information may be produced, given that part of it is expected to leak out to uninformed investors through price movements driven by information-based trades.

In an extension of our benchmark model we allow for smart-money trading both on the organized exchange and on OTC markets. When financiers have a choice of becoming either informed traders on the exchange or informed dealers in an OTC market, we show that in equilibrium OTC markets are always too large relative to the organized exchange. The reason is that substitution of informed trading on a transparent exchange for trading on opaque OTC markets results in worse terms of trade for entrepreneurs with good assets. Therefore, to maintain the same origination incentives of good assets by entrepreneurs a larger informed

34

financial sector is required. Given that information rents are bigger in opaque OTC markets, financiers' private incentives are to switch trading to OTC markets even if this tends to undermine entrepreneurs' ex-ante incentives to originate good assets, which explains why OTC markets are too large in equilibrium.

Informed dealers profit from the opaqueness of OTC transactions and this is one reason why broker-dealers have generally resisted the transfer of trading of the most standardized OTC contracts onto organized platforms, as required by the Dodd-Frank Act of 2010[21]. This is also why the largest Wall Street firms are so intent on avoiding disclosure of prices and fees in the new exchanges set up in response to the Dodd-Frank Act.[22] Interestingly, in a heterogeneous world, firms with a high probability of generating good projects also benefit (*ex-ante*) from the option of trading in opaque markets. It is, thus, not surprising that some firms have also been keen to keep OTC markets in their present form.[23] All in all, we therefore expect that a first line of defense by the financial industry to the new regulations required under the Dodd-Frank Act is likely to be to over-customize derivatives contracts and to offer fewer standardized, plain-vanilla, contracts (which will be required to trade on organized exchanges); the second line of defense will be to set up clearinghouses that maintain opacity and do not require disclosure of quotes; and a third line will be to ensure that the operation of clearinghouses remains under the control of the main dealers.

---

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

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

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

REFERENCES

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

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

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

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

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

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

Bolton, Patrick, Tano Santos and Jose Scheinkman (2012) "Shadow Finance,"to appear in "*Rethinking Finance*," A. Blinder, A. Lo and R. Solow (eds).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Philippon, Thomas (2012) "Finance vs. Walmart: Why are Financial services so Expensive?,"to appear in "*Rethinking Finance*," A. Blinder, A. Lo and R. Solow (eds).

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

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

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

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

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

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

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

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

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

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

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

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

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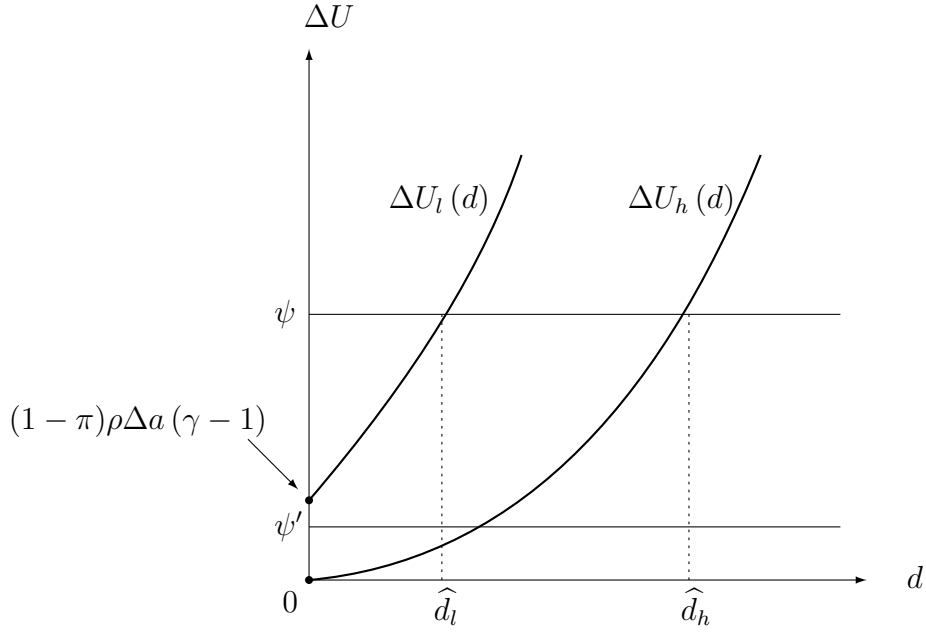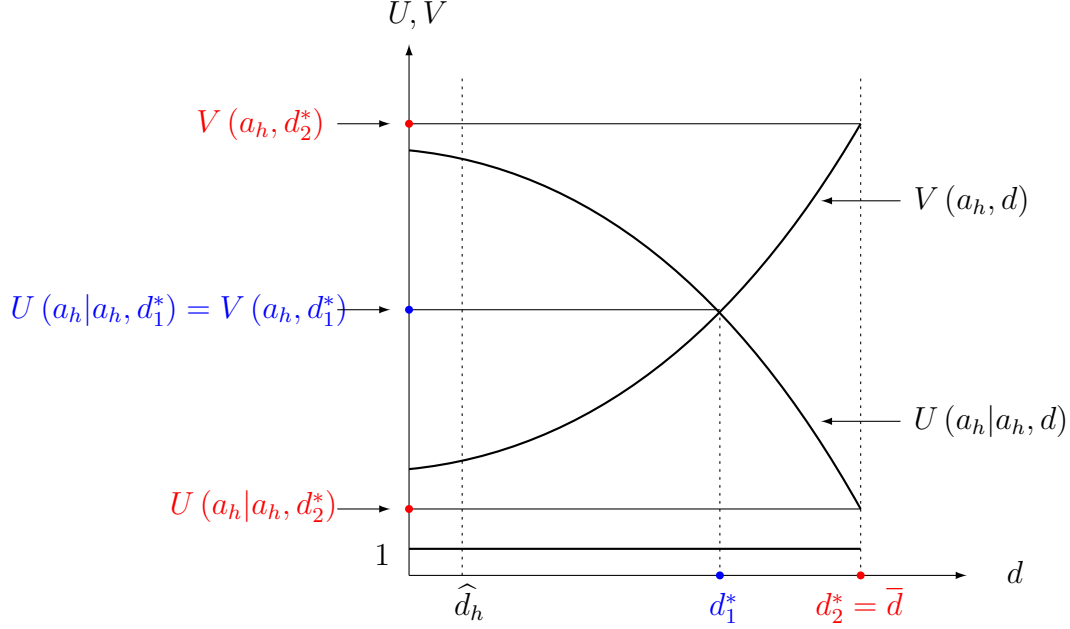

Figure 2: **Incentive compatibility.**
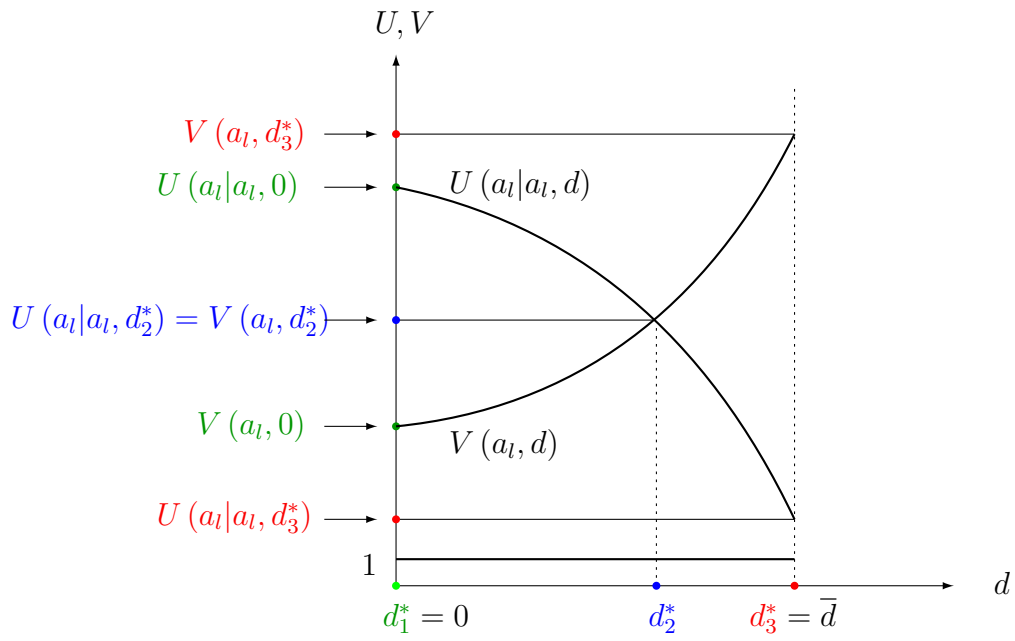

Figure 3: **High effort equilibria.**

40

Figure 4: **Low effort equilibria.**

APPENDIX

Not for publication

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

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

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

Proof of Proposition 4. Note that

$$\frac{\partial m}{\partial d} = \frac{(1-\pi)}{\pi a (1-d)^2} > 0 \qquad and \qquad \frac{\partial^2 m}{\partial d^2} = \left(\frac{2}{1-d}\right)\frac{\partial m}{\partial d} > 0, \tag{31}$$

and

$$\frac{\partial p}{\partial d} = \left[\frac{a\rho(1-a)(1-\gamma)}{[a(1-m)+(1-a)]^2}\right]\frac{\partial m}{\partial d} < 0 \quad as \quad \gamma > 1. \tag{32}$$

Finally,

$$\frac{\partial^2 p}{\partial d^2} = \frac{a\rho(1-a)(1-\gamma)}{[a(1-m)+(1-a)]^2}\left[\frac{\partial^2 m}{\partial d^2} + \frac{2}{a(1-m)+(1-a)}\left(\frac{\partial m}{\partial d}\right)^2\right] < 0. \tag{33}$$

Expressions (31), (32), and (33) are used throughout. $\square$

Proof of Proposition 3. From (9),

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

42

and

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

given (32) and (33), which establishes (b).

As for the utility of the entrepreneur, (8), note that

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

It can be shown that

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

and hence

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

and thus we can write

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

Finally,

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

which proves (a). □

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

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

$$U_{hl}\left(d\right) = p_h\left(d\right) + a_l m_h\left(d\right)\left(\gamma\rho - p_h\left(d\right)\right)\left(\pi\kappa + \left(1 - \pi\right)\right). \tag{34}$$

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

$$U_{lh}\left(d\right) = -\psi + \pi\left[p_l(d) + a_h m_l(d)\kappa\left(\gamma\omega\rho - p_l(d)\right)\right] + \left(1 - \pi\right)\omega\rho\left[1 + a_h\left(\gamma - 1\right)\right] \tag{35}$$

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

$$U^{sell}\left(a_l|a_h, d, no - liq.\right) \qquad and \qquad U^{no-sell}\left(a_l|a_h, d, no - liq.\right), \tag{36}$$

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

43

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

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

First, notice that

$$
\begin{aligned}
U^{sell} (a_l | a_h, d, no - liq.) &= a_l m_h(d) \gamma \rho + (1 - a_l m_h(d)) p_h(d) \qquad (37) \\
&= p_h(d) + a_l m_h(d) (\gamma \rho - p_h(d)),
\end{aligned}
$$

where the functions $p_h(d)$ and $m_h(d)$ were given by (**??**) and (**??**), respectively, when $a = a_h$. The first term in (37) is the payoff, conditional on having a good project and receiving a bid from a dealer, and event with probability $a_l m_h(d)$, in which case the entrepreneur rejects the bid and carries the project to maturity and obtains, $\gamma \rho$, as recall that he is not subject to the liquidity shock. The second term is the payoff when he does not receive a bid but sells anyway. Since $p_h = p_h(m(d))$ we may consider the function

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

Notice that

$$U^{no-sell} (a_l | a_h, d, no - liq.) = \rho [1 + a_l (\gamma - 1)] = f(1). \qquad (38)$$

Further,

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

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

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

and hence

$$U_{hl} (d) = \pi [p_h(d) + a_l m_h(d) \kappa (\gamma \rho - p_h(d))] + (1 - \pi) U^{sell} (a_l | a_h, d, no - liq.), \qquad (39)$$

which after some manipulations yields (34).

(b) We show that

$$U^{sell} (a_h | a_l, d, no - liq.) \leq U^{no-sell} (a_h | a_l, d, no - liq.) \quad \text{for all} \quad d. \qquad (40)$$

First notice that,

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

44

Second, notice that

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

and, as before, define

$$g(m) = p_l(m) + a_h m(\gamma\rho - p_l(m)).$$

Notice that

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

Finally, we can show that

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

Thus for every m, $g(m) \leq U^{no-sell}(a_h|a_l, d, no - liq.)$, establishing (40). Thus

$$U(a_h|a_l, d) = -\psi + \pi[p_l(d) + a_h m_l(d)\kappa(\gamma\omega\rho - p_l(d))] + (1-\pi)U^{no-sell}(a_h|a_l, d, no - liq.). \quad (41)$$

Trivial manipulations of (41) yield (35). $\qquad\square$

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

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

by Proposition 4. Similarly notice that

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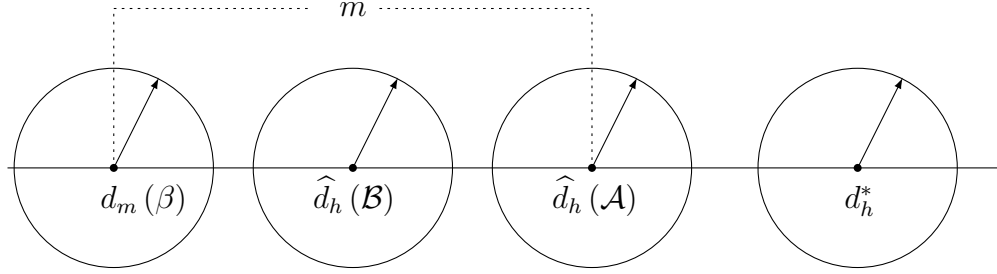
(b)

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

as

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

by Proposition 4. $\qquad\square$

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

Proof of Proposition 7. First fix the vector $\mathcal{A}$, and make φ left-continuous by setting $\varphi(\bar{d}) = \lim_{d\uparrow\bar{d}}\varphi(d)$. Consider the function

$$L(d,\beta,\mathcal{A}) := \varphi(d) + \beta - 1 - (1-\pi)(1-\kappa)(\rho\gamma - p(a_h,d)) + U(a_h|a_h,d) \qquad (42)$$

Since L is smooth, and $\frac{\partial L}{\partial \beta} = 1$, the Transversality Theorem guarantees that for almost every β the positive solutions d to equation (42) satisfy $\frac{\partial L}{\partial d}(d,\beta,\mathcal{A}) \neq 0$ and in particular these solutions are isolated. Choose one such β with $0 < \beta < \epsilon$, and such that $L(\bar{d},\beta,\mathcal{A}) \neq 0$. If $d = \bar{d}$ is an equilibrium with high effort then necessarily $L(\bar{d},\beta,\mathcal{A}) < 0$ and if $d < \bar{d}$ is an equilibrium with high effort then necessarily $L(d,\beta,\mathcal{A}) = 0$, since the marginal dealer in such equilibria must be indifferent between becoming a dealer or an entrepreneur. Furthermore, since $L(\bar{d},\beta,\mathcal{A}) \neq 0$, there are only a finite number of zeros of $L(d,\beta,\mathcal{A})$. If there are solutions to $L(d,\beta,\mathcal{A}) = 0$ that are strictly less than $\widehat{d}_h(\mathcal{A})$, let $d_m(\beta)$ be the largest such solution and $m = \widehat{d}_h(\mathcal{A}) - d_m(\beta)$ (see the figure above.) Otherwise set $m = \widehat{d}_h(\mathcal{A})$. Since high effort is constrained efficient, $\Delta U_h(\widehat{d}_h(\mathcal{A})) = \psi$, and the proof of Proposition 5 establishes that $\widehat{d}_h(\cdot)$ is differentiable. Choose $\mathcal{B} = (a_h, a'_l, \gamma, \rho, \kappa, \pi, \psi, \bar{d})$ with $a_l - \epsilon < a'_l < a_l$ (notice we keep all other parameter values unchanged). Optimality of high effort is maintained, but ΔU_h shifts up and thus $\widehat{d}_h(\mathcal{B}) < \widehat{d}_h(\mathcal{A})$. Since $\widehat{d}_h(\cdot)$ is differentiable, by choosing a'_l close enough to a_l we guarantee that $\widehat{d}_h(\mathcal{B}) > \widehat{d}_h(\mathcal{A}) - m$. Further, the differentiability of $\widehat{d}_h(\cdot)$ insures that we may choose $0 < \eta < m$ and neighborhood $\mathcal{N}$ of $\mathcal{B}$ such that for each $\mathcal{B}' \in \mathcal{N}$, $\eta \leq \widehat{d}_h(\mathcal{A}) - \widehat{d}_h(\mathcal{B}') \leq m - \eta$. Notice that the expression in (42) remains unchanged and thus the set of solutions to $L(d,\beta,\mathcal{A}) = 0$ is the same as the set of solutions to $L(d,\beta,\mathcal{B}) = 0$, and $L(\bar{d},\beta,\mathcal{A}) < 0$ if and only if $L(\bar{d},\beta,\mathcal{B}) < 0$. Now, since at any zero of $L((d,\beta,\mathcal{A}), \frac{\partial L}{\partial d}(d,\beta,\mathcal{A}) \neq 0$ and there are only a finite number of these zeros, we may choose δ

46

such that for $|\beta' - \beta| < \delta$ and $\mathcal{B}'$ in an open ball $\mathcal{O} \subset \mathcal{N}$, such that if $L(d(\beta', \mathcal{B}'), \beta, \mathcal{B}') = 0$, then there exists a $d(\beta, \mathcal{A})$ with $L(d(\beta, \mathcal{A}), \beta, \mathcal{A}) = 0$ such that $|d(\beta', \mathcal{B}') - d(\beta, \mathcal{A})| < \frac{\eta}{2}$. If $d(\beta, \mathcal{A}) \geq \widehat{d}_h(\mathcal{A})$ then, $\widehat{d}_h(\mathcal{B}') < d(\beta', \mathcal{B}') - \frac{\eta}{2}$. On the other hand, if $d(\beta, \mathcal{A}) < \widehat{d}_h(\mathcal{A})$ then $\widehat{d}_h(\mathcal{A}) - d(\beta, \mathcal{A}) \geq m$ and, $\widehat{d}_h(\mathcal{B}') > d(\beta', \mathcal{B}') + \frac{\eta}{2}$. In any case, $\widehat{d}_h(\mathcal{B}')$ does not solve equation (42) for any β' with $|\beta' - \beta| < \delta$. Since any equilibrium $d_h^*(\beta', \mathcal{B}') < \bar{d}$ must be a solution to equation (42) that satisfies $d_h^*(\beta', \mathcal{B}') \geq \widehat{d}_h(\mathcal{B}')$, we must have that $d_h^*(\beta', \mathcal{B}') > \widehat{d}_h(\mathcal{B}')$, for every $\mathcal{B}' \in \mathcal{O}$ and $|\beta' - \beta| < \delta$. Thus, for these parameters, all equilibria with $d < \bar{d}$ are inefficient and any high effort equilibrium features too many dealers. In addition, if there is an equilibrium at $\bar{d}$ since $\bar{d} \geq \widehat{d}_h(\mathcal{A}) > \widehat{d}_h(\mathcal{B}')$, this equilibrium is also inefficient. $\qquad\square$

Proof of Proposition 8. This follows immediately from the monotonicity of the expression in (20) . $\qquad\square$