Is the financial sector too big?*

Patrick Bolton, Columbia University
Tano Santos, Columbia University
Jose A. Scheinkman, Princeton University

May 28, 2010

Abstract

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

*Very preliminary draft prepared for the *6th Banco de Portugal Conference on Monetary Economics* held at Lisboa on June 10-11 2010. Please do not quote.

1 Introduction

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

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

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

In his survey of the literature on financial development and growth, Levine (2005) synthesizes existing theories of the role of the financial industry into five broad functions: 1) information production about investment opportunities and allocation of capital; 2) mobilization and pooling of household savings; 3) monitoring of investments and performance; 4) financing of trade and consumption; 5) provision of liquidity, facilitation of secondary market trading, diversification, and risk management. As he highlights, most of the models of the financial industry focus on the first three functions, and if anything, conclude that from a social efficiency standpoint the financial sector is too small. That is, if lending and capital provision by the financial industry were to grow, output and welfare would also grow. In other words, most of these models conclude that due to asymmetries of information, and incentive or contract enforceability constraints, there is underinvestment in equilibrium and *financial underdevelopment*. These models provide theoretical underpinnings to the empirical findings from cross-country regressions that financial development leads growth.

In contrast to this literature, our model mainly emphasizes the fifth function of the financial industry in Levine's list: secondary market trading and liquidity provision. In addition, where the finance and growth literature only distinguishes between bank-based and market-based systems (e.g. Allen and Gale, 2000), a key departure of our model is the distinction we draw between *organized exchanges* and over-the-counter (*OTC*) markets.

As is illustrated in Figure 0 below, which plots the evolution of wages in the U.S. banking, insurance and 'other finance' sectors from the great depression onwards, the key growth in remuneration in the financial industry at large has taken place in investment banking, derivatives trading and OTC markets. Thus, in order to understand whether the financial industry extracts too high rents and as a result attracts too much talent, one needs to focus on why remuneration in these markets is so high.

In our model, secondary market trading requires information about underlying asset quality and valuation skills. When an entrepreneur is looking to sell his firm in the secondary market the buyer must be able to determine the value of the firm that is up for sale. This is where the *young talent* employed in the financial industry manifests itself. *Informed dealers* in the OTC market are able to determine the value of assets for sale and can offer to buy the most valuable assets from entrepreneurs. By identifying the most valuable assets and by offering more attractive terms for those assets than are available in the organized market, informed dealers in the OTC market provide incentives to entrepreneurs to originate good assets. However, the central efficiency question for agents' occupational choices between the financial and real sectors is what share of the incremental value of these good assets dealers get to appropriate. Valuation skills in reality and in our model are costly to acquire and generally scarce. This is why not all asset sales can take place in the OTC market. Those

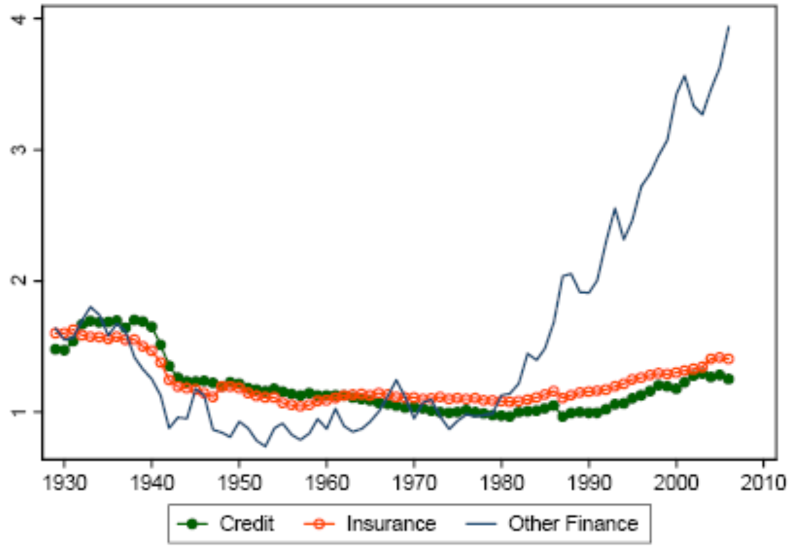

Figure 0: Wages in the financial sector relative to non farm private sector. Figure 2-B in T. Philippon and A. Reshef, "Wages and Human Capital in the US Financial Industry: 1926-2006," NYU working paper, December 2008.

assets that cannot be absorbed by the OTC market end up on the organized exchange. The relative scarcity of *informed capital* in the OTC market is a key determinant of the size of the information rents that are extracted by the financial sector in equilibrium.

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

In other words, as in Rothschild and Stiglitz (1976), informed dealers in the OTC market are able to *cream-skim* the best assets and thus extract an informational rent. This cream-skimming activity of informed dealers imposes a negative price externality on the organized market, as uninformed investors operating in this market understand that they only get to buy an adversely selected pool of assets. This negative price externality in turn weakens the bargaining position of entrepreneurs selling good assets in the OTC market, as their threat-point of selling the asset in the organized market becomes less attractive.

This is why in our model rent-extraction by informed dealers actually increases as the OTC market expands. This is also the key mechanism which: *i)* gives rise to an excessively large financial industry in equilibrium; *ii)* is why there is excessive information rent-extraction in the OTC market; and, *iii)* is why the financial sector–in particular the OTC market–attracts too much talent. All the human capital invested in young talent to train an informed dealer in the end mainly serves to extract informational rents rather than create social surplus by allowing entrepreneurs who originate valuable assets to realize a fair value for those assets when they sell them in secondary markets. Our model thus helps explain how excessive rent extraction and entry into the financial industry can be an *equilibrium outcome*, and why competition for rents doesn't eliminate excessive rent extraction.

The structure of the financial industry, combining an informal OTC market and an organized exchange is a key feature of our theory. Unlike models of informed trading in the tradition of Grossman and Stiglitz (1980) in our model dealer information in the OTC market is asset specific and cannot be reflected in a market price, as each transaction is an undisclosed bilateral deal between the dealer and the seller of the asset. Therefore, when more dealers compete in the OTC market, this does not result in more information being transmitted through prices. On the contrary, more competition by informed dealers simply results in more cream-skimming and more information rent extraction. Our highly stylized model of the financial industry can be seen as an *allegory* of a general phenomenon in the financial industry, where informed parties have an incentive to trade and remove themselves from organized markets. This is not just true for derivatives and swaps, which are mostly traded in OTC markets, but also for secondary stock markets, where trading by institutions often takes place in an informal 'upstairs market' or more recently in so called 'dark pools'. Similarly, for primary markets, private equity funding or private placements have the same cannibalizing effect on organized exchanges, of removing the better and hardest to value assets from uninformed investors' reach.

Our paper contributes to a small literature on the optimal size of the financial industry. An early theory by Murphy, Shleifer and Vishny (1991) (see also Baumol, 1990) builds on the idea of increasing returns to ability and rent seeking to show that in a two-sector model there may be inefficient equilibrium occupational outcomes, where too much talent enters one market since the marginal private returns from talent exceed the social returns. More recently, Philippon (2008) has proposed an occupational choice model where agents can choose to become workers, financiers or entrepreneurs. The latter originate projects which have a higher social than private value, and need to obtain funding from financiers. In general, as social and private returns from investment diverge it is optimal in his model to subsidize entrepreneurship. Neither the Murphy et al. (1991) nor the Philippon (2008) models distinguish between organized exchanges and OTC markets in the financial sector, nor do they allow for excessive informational rent extraction through cream-skimming. In independent work Glode, Green and Lowery (2010) also model the idea of excessive investment in information as a way of strengthening a party's bargaining power. However, Glode et al. (2010) do not consider the occupational choice question of whether too much young talent is attracted towards the financial industry. Finally, Lagos, Rocheteau and Weill (2009) propose a model of the OTC market that has common elements to ours. However, their focus is on the liquidity of this market and they do not address issues of cream-skimming or occupational choice.

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

2 The model

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

2.1 Agents.

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

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

Type 2 agents are the *active population.* Each type 2 agent can work either as a (self-employed) *entrepreneur* in the real sector, or as a *dealer* in the financial sector. Type 2 agents make an occupational choice decision in period 0 to which they are commited to for the remainder of their life.

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

We simplify the model without much loss in generality by assuming that all type 2 agents start in period 1 with the same unit endowment, $\omega = 1$, have the same preferences over consumption, face the same idiosyncratic liquidity shocks, and are equally able entrepreneurs. Type 2 agents can only differ in their ability to become well-informed dealers. Specifically, we represent the mass of type 2 agents by the unit interval $[0, 1]$ and order these agents $d \in [0, 1]$ in the increasing order of the costs they face of acquiring the human capital to become well informed dealers: $\varphi(d)$. We then assume that $\varphi(d)$ is non-decreasing and that

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

where $\overline{d} < 1$.

In all other respects, type 2 agents are identical:

1. They face the same *i.i.d.* liquidity shocks and value consumption only in period 1 with probability π and only in period 2 with probability $(1 - \pi)$. Their liquidity preferences,

whether they choose to become entrepreneurs or dealers, are thus represented by the utility function

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

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

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

$$(\gamma - 1)\rho(a_h - a_l) \equiv (\gamma - 1)\rho\Delta a > \psi.$$

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

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

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

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

endowment, together with any collateralized borrowing, to purchase assets for sale by impatient entrepreneurs. If they learn that they are impatient they simply consume their unit endowment.

2.2 Financial Markets

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

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

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

Thus, in period 1 a dominant strategy for impatient entrepreneurs is to attempt to first approach a dealer. They understand that with probability $a \in \{a_l, a_h\}$ the underlying value of their asset is high, in which case they are able to negotiate a sale with a dealer at price $p^d > p$ with probability $m \in [0, 1]$. If they are not able to sell their asset for price p^d to a dealer, entrepreneurs have no choice but to turn to the organized market in which they can sell their asset for price p.

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

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

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

The price p^d at which a sale is negotiated between a dealer and an entrepreneur is the outcome of a bilateral bargaining game (under symmetric information). The price p^d has to exceed the *status-quo* price p in the organized market at which the entrepreneur can always sell his asset. Similarly, the dealer cannot be worse off than under no trade, when his payoff is 1, so that the price cannot be greater than the value of the asset $\gamma\rho$. We take the solution to this bargaining game to be given by the *Asymmetric Nash Bargaining Solution*, where the dealer has bargaining power $(1-\kappa)$ and the entrepreneur has bargaining power κ (see Nash, 1950, 1953).[3] 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

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

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

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). [4]

The price p^d may be higher than 1, the dealer's endowment. In that case the dealer needs to borrow the difference $(p^d - 1)$ against the asset to be acquired. As long as this difference does not exceed ρ, the dealer will not be financially constrained. For simplicity, we shall restrict attention to parameter values for which the dealer is not financially constrained. We provide a condition below that ensures that this is the case. [5]

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

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

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

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

[4] 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 π denotes the probability of an entrepreneur subsequently matching with another informed dealer then Binmore, Rubinstein and Wolinsky show that $\kappa = \pi$.

[5] 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.

Note that p is decreasing in m, from the highest price $p = \rho[a(\gamma - 1) + 1]$ when $m = 0$ to the lowest price $p = \rho$ when $m = 1$.

2.3 Timing

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

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

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

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

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

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

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

7. As for patient entrepreneurs, we will show that along the equilibrium path they hold on to the asset they originated in period 0 until maturity and then consume the asset's output. Their expected consumption is then $\rho[a(\gamma - 1) + 1]$. In a symmetric equilibrium where all entrepreneurs choose $a = a_h$, we then need to check that the double deviation, where a single entrepreneur chooses $a = a_l$ in period 0 and sells his asset in period 1 even if he learns that he is patient is not profitable.

2.4 Discussion and Parameter Restrictions

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

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

We have introduced ex-ante heterogeneity among type 2 agents only in the form of different non-pecuniary costs in acquiring information to become a dealer. We could also, or alternatively, have introduced heterogeneity in the costs of becoming an entrepreneur. Nothing substantive would be added by introducing this other form of heterogeneity. We would then simply order type 2 agents in their increasing *comparative advantage* of becoming dealers and proceed with the analysis as in our current model. For simplicity we have therefore suppressed this added form of heterogeneity.

The reader may wonder why we introduce any ex ante heterogeneity among type 2 agents at all? It turns out that the greater generality of the model with ex-ante heterogeneous type 2 agents actually gives rise to an analytically more tractable model. Indeed, with ex-ante identical type 2 agents, all these agents would have to be indifferent between becoming dealers or entrepreneurs in equilibrium and supporting such an equilibrium would require that type 2 agents randomize their choice between their two occupations. Characterizing such a mixed-strategy equilibrium would be analytically more involved and would not give rise to a simpler analysis.[6]

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

$$m(a,d) = \frac{d(1-\pi)}{a(1-d)\pi} < 1. \tag{6}$$

We thus require that π is sufficiently large, and/or $\overline{d}$ sufficiently small so that $d < \overline{d}$ remains small. 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 $m > 1$, this does not seem to be the empirically plausible parameter region. When $m > 1$ 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 $m > 1$, this is unlikely to happen in equilibrium, as dealers would then compete

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

away their informational rents to the point where they would not be able to recoup even their relatively low investment in dealer skills $\varphi(d)$.

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

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

We provide an assumption on parameter values guaranteeing that this inequality holds in the appendix. A simple sufficient condition for this inequality is that the equilibrium measure of informed dealers remains small, which is the case if $\overline{d}$ is small.

There are basically two non-trivial optimization problems in our model. The first is the occupational choice problem of type 2 agents. The second is the effort choice problem of entrepreneurs in period 0. As the latter problem is somewhat more involved we begin our analysis with this moral hazard problem.

3 Entrepreneur Moral Hazard

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

For simplicity, we restrict attention to symmetric equilibria in which all entrepreneurs choose the same effort in period 0. A necessary condition for any symmetric equilibrium then is that it is *incentive compatible* for entrepreneurs to choose the equilibrium effort. In other words, entrepreneurs must (weakly) prefer to choose the equilibrium action than to deviate to the other action.

Consider first incentive compatibility in the high effort equilibrium, where all entrepreneurs choose a_h. In the high effort equilibrium, the matching probability in period 1 and market prices are denoted by respectively p_h and m_h. An entrepreneur's expected payoff in period 0 when choosing effort a_h in the high effort equilibrium is then given by:

$$\begin{aligned} U^h &= -\psi + \pi \left\{ a_h m_h \left(\kappa\omega\gamma\rho + (1-\kappa) p_h \right) + (1 - a_h m_h)\, p_h \right\} \\ &+ (1-\pi)\omega\rho \left[1 + a_h (\gamma - 1) \right] \\ &= -\psi + \pi \left[p_h + a_h m_h \kappa \left(\gamma\omega\rho - p_h \right) \right] + (1-\pi)\omega\rho \left[1 + a_h (\gamma - 1) \right] \end{aligned} \tag{7}$$

where,

$$p_h = \frac{\rho[a_h(1-m_h)\gamma + (1-a_h)]}{1 - a_h m_h}$$

(for notational convenience we have suppressed the dependence of p_h and m_h on d^h).

Suppose now that an entrepreneur chooses to *deviate* in period 0 by choosing the low effort a_l. It can be shown that in this case that it is optimal for such an entrepreneur to put his asset for sale in the OTC market when he is not hit by a liquidity shock. This entrepreneur would not accept the offer by the dealer as he would know, upon getting the offer, that he possesses a good project. However if he does not receive an offer he would prefer to sell his asset in the uninformed exchange to holding on to it. One can show that the payoff of an entrepreneur that deviates to the low effort is given by,

$$U^{hl} = p_h + a_l m_h \left(\gamma\omega\rho - p_h \right) \left(\pi\kappa + (1-\pi) \right) \tag{8}$$

(the superscript, hl, refers to the payoff from a *deviation* from a_h to a_l).

Incentive compatibility in the high effort equilibrium then requires that $U^h \geq U^{hl}$, or, denoting by $\Delta U^h(d^h)$ the difference in expected *monetary payoff* from the high versus the low effort, that when the measure of dealers is d^h:

$$\begin{aligned} \Delta U^h \left(d^h \right) &= \pi \Delta a m_h \kappa \left(\omega\gamma\rho - p_h \right) \\ &+ (1-\pi) \left[\omega\rho \left(1 + a_h (\gamma - 1) \right) - \left(p_h + a_l m_h \left(\gamma\omega\rho - p_h \right) \right) \right] \geq \psi \end{aligned} \tag{9}$$

Now consider incentive compatibility in the low effort equilibrium, where all entrepreneurs choose a_l. In this equilibrium, market prices and the matching probability in period 1 are given by p_l and m_l, respectively. An entrepreneur's expected payoff in period 0 when choosing the low effort is then:

$$\begin{aligned} U^l &= \pi \left\{ a_l m_l \left(\kappa\omega\gamma\rho + (1-\kappa) p_l \right) + \left(1 - a_h m_l\right) p_l \right\} \\ &+ (1-\pi)\omega\rho \left[1 + a_l (\gamma - 1)\right] \\ &= \pi \left[p_l + a_l m_l \kappa \left(\gamma\omega\rho - p_l\right)\right] + (1-\pi)\omega\rho \left[1 + a_l (\gamma - 1)\right] \end{aligned} \tag{10}$$

where,

$$p_l = \frac{\rho[a_l(1-m_l)\gamma + (1-a_l)]}{1 - a_l m_l}.$$

An entrepreneur who chooses to *deviate* from this equilibrium in period 0 by choosing the high effort a_h, in which case he is better off holding on to his asset until period 2, unless he is hit by a liquidity shock in period 1, then gets:

$$U^{lh} = -\psi + \pi \left[p_l + a_h m_l \kappa \left(\gamma\omega\rho - p_l\right)\right] + (1-\pi)\omega\rho \left[1 + a_h (\gamma - 1)\right] \tag{11}$$

Incentive compatibility in the low effort equilibrium again requires that $U^l \geq U^{lh}$, or that the difference in expected *monetary payoff* from the high versus low effort, $\Delta U^l(d^l)$, is such that:

$$\Delta U^l \left(d^l\right) = \pi \Delta a m_l \kappa \left(\gamma\omega\rho - p_l\right) + (1-\pi)\omega\rho\Delta a (\gamma - 1) \leq \psi \tag{12}$$

As will become clear, it is helpful for the characterization of equilibrium that follows to consider the functions $\Delta U^h(d)$ and $\Delta U^l(d)$, which give the difference in expected *monetary payoff* from a deviation away from the equilibrium action a^* as a function of an exogenously given measure of informed dealers d (for respectively the high and low-origination effort candidate equilibria).

4 Occupational Choice and Equilibrium

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

There are two types of equilibria, which may co-exist. One is a **low-origination-effort equilibrium**, in which all entrepreneurs choose $a = a_l$, and where the financial sector is **small** (a small measure of type 2 agents choose to become dealers). The other is a **high-origination-effort equilibrium**, in which all entrepreneurs choose $a = a_h$, and where the financial sector is **large;** that is, a sufficiently large measure of type 2 agents choose to become dealers that impatient entrepreneurs with good assets in period 1 are likely enough to match with a dealer. 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.

4.1 Equilibrium borrowing and asset trading in period 1

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

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

Proof. Consider first an impatient entrepreneur. By selling his asset in the organized market he is able to obtain at least p, which is higher than the maximum amount ρ he can borrow against the asset. Therefore, an impatient entrepreneur strictly prefers to sell his assets than to borrow. As for a patient entrepreneur, since he strictly prefers to consume in period 2 he cannot gain by borrowing and consuming in period 1. He also cannot gain (strictly) from borrowing and investing the proceeds from the loan in either the organized or OTC markets. A patient entrepreneur is no different as an investor than an uninformed type 1 agent, and therefore earns the same zero net returns in equilibrium as type 1 agents. Finally, consider an impatient dealer. Such a dealer can only borrow against an asset he has acquired in either the OTC or organized market. Moreover, he can only gain from acquiring and borrowing against an asset if he is able to resell the asset for a profit. But this is not possible in either market at equilibrium prices p and p^d. ■

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

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

Lemma 2 *A* patient *entrepreneur, who follows the equilibrium action $a^* \in \{a_l, a_h\}$ in period* 0 *(weakly) prefers not to put up his asset for sale in period* 1*.*

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

In period 1 all *impatient* entrepreneurs put their assets up for sale. They first approach a dealer in the OTC market and if they are unable to generate an offer from a dealer they sell their asset in the organized market. As we have shown above, in any equilibrium (a^*, d^*) with a measure d^* of dealers and where all entrepreneurs have chosen the same origination action

a^* in period 0, the equilibrium matching probability of an impatient entrepreneur (holding a good asset) with a patient dealer in period 1 is given by:

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

And, given this equilibrium matching probability, the equilibrium price for an asset in the organized market is

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

so that the price in the OTC market is $p^d = \kappa\gamma\rho + (1-\kappa)p^*$.

Our next result establishes a central property of our equilibrium and shows how the equilibrium price varies with a^* and d^*.

Proposition 3 *(a) $p(a_l, d) < p(a_h, d)$ for all $d \geq 0$ and (b) $p_d(a^*, .) < 0$ and $p_{dd}(a^*, .) < 0$ for all $a^* \in \{a_l, a_h\}$, so that the equilibrium price in the organized market is a decreasing and concave function of the measure of dealers.*

Proof. (a) is immediate and (b) follows from differentiation of p with respect to d for a given a and the definition of m.

The first part of the proposition is not surprising. When entrepreneurs are expected to originate higher quality assets ($a^* = a_h$) in period 0 then for a given measure of dealers d the probability of an entrepreneur (holding a good asset) matching with a dealer in the OTC market, $m^*(a_h, d)$ is lower, so that the proportion of good assets being sold in the organized market is higher. With a higher proportion of good assets expected to be sold in the organized market the equilibrium price of assets in that market is higher. The second part of the proposition is more surprising and is the central mechanism in our model. When the mass of informed dealers increases, the probability of an entrepreneur (holding a good asset) matching with a dealer in the OTC market, $m^*(a^*, d^*)$ is higher, and consequently the proportion of good assets being sold in the organized market is lower. This is why the price in the organized market declines when more deals get executed in the OTC market. In other words, dealers in the OTC market *cream-skim* the good assets and thereby impose a *negative externality* on the organized market.

Importantly, this negative externality on the organized market results in an improvement in the terms at which dealers get to purchase good assets (at price $p^d = \kappa\gamma\rho + (1-\kappa)p^*$). When p^* declines, entrepreneurs negotiating with dealers in the OTC market face worse outside options and therefore are willing to make more price concessions. Thus, when more dealers are present in the OTC market, far from dissipating their informational rent, they are able to increase the fraction of that rent they can extract in equilibrium. Thus, an important prediction of our model is that dealers obtain a higher payoff as the OTC market grows.

4.2 Equilibrium Payoffs in period 0

We are now in a position to determine equilibrium payoffs for dealers and entrepreneurs in period 0. Suppose that in the (symmetric) equilibrium a fraction d of type 2 agents decide to become dealers and a fraction $(1-d)$ decide to become entrepreneurs. Suppose, in addition, that all entrepreneurs choose the equilibrium action a^*. Then, an entrepreneur's expected utility in period 0 is given by

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

(where, for notational convenience we have suppressed the dependence of m^* and p^* on d and a^*).

As is easy to verify, an entrepreneur's payoff in period 0 is decreasing in the equilibrium measure of dealers d: $U_d(a^*|a^*,.) < 0$. Note also that $U_{dd}(a^*|a^*,.) < 0$, so that an entrepreneur's ex-ante payoff is a decreasing, concave, function of the measure of dealers d. In other words, far from benefiting from a marginal increase in the financial sector, entrepreneurs actually are made worse off. This result captures in a nutshell the populist sentiment of *Main street* towards *Wall street*.

The expected utility of dealer $\widetilde{d} \le d$ in period 0 in this same equilibrium is given by

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

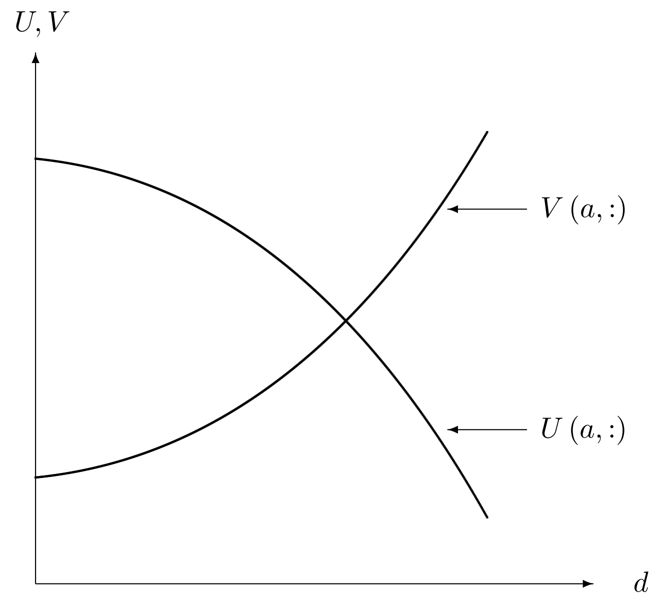

Figure 1: **Payoff functions.** Utility functions of the entrepreneur and a *given* dealer as a function of the measure of dealers, d

Again, it is easy to verify that $V_d(\widetilde{d} \mid a^*, .) > 0$ and also that $V_{dd}(d \mid a^*, .) > 0$. In words, dealers are better off operating in bigger OTC markets, in which more good deals get *skimmed off* from organized markets. Another interesting observation is that dealers also prefer dealing in market equilibria with low quality origination of assets (it is immediate to verify that $V(\widetilde{d} \mid a_h, d) < V(\widetilde{d} \mid a_l, d)$). The reason is that the negative cream-skimming externality of dealers on organized markets is larger when a low fraction of good assets is originated.

In sum, if it were not for the positive incentive effects of cream-skimming in OTC markets on entrepreneurs, informed dealers would mostly play a parasitical role in our economy. They would enrich themselves by helping entrepreneurs with good assets get a better price, but they would not create any net social surplus.

4.3 Equilibrium Size of the Financial Sector

We now turn to a key question we are interested in: what is the equilibrium size of the financial sector? In our model this question boils down to determining the equilibrium measure of dealers d^*. As we have already highlighted, there may be two types of equilibria, each with an associated size of the OTC market. One type of equilibrium is the low-origination-effort equilibrium, in which all entrepreneurs choose $a = a_l$. As we show below, this equilibrium is associated with a small measure of dealers d^l. The other type of equilibrium is the high-origination-effort equilibrium, in which all entrepreneurs choose $a = a_h$. This equilibrium has an associated large financial sector with a large measure of dealers d^h.

The broad intuition for why low origination effort is coupled with a small measure of dealers in equilibrium is that with a smaller supply of good assets brought to the OTC market a smaller mass of informed dealers is needed to trade these assets. When one takes into account all the general equilibrium effects, however, this intuition is somewhat misleading. A first complicating factor to note is that a smaller mass of dealers d^l also means a higher mass of entrepreneurs $(1 - d^l)$. Even if entrepreneurs originate a lower fraction a_l of good assets per capita, it is then not entirely obvious *a priori*, that in aggregate the supply of good assets is smaller. As already noted, a second complicating factor is that, other things equal, dealers prefer dealing in an economy with low origination of good assets. Again, it is then not immediately obvious that d^l will not be larger than d^h.

Now, a necessary condition for the existence of either type of equilibrium is that it is incentive compatible for entrepreneurs under each equilibrium to choose the prescribed equilibrium action. This means that in the high-origination-effort equilibrium, $\Delta U^h(d^h)$ (the difference in expected monetary payoff from the high versus the low effort) must satisfy condition (9), and in the low-origination-effort equilibrium, $\Delta U^l(d^l)$ must satisfy condition (12). The next lemma establishes a clear and critical ranking of the functions $\Delta U^h(d)$ and $\Delta U^l(d)$, which underlies the result we prove next that $d^l < d^h$.

Lemma 4 *(a) $\Delta U^h(d)$ and $\Delta U^l(d)$ are both increasing functions of d and (b) $\Delta U^h(d) < \Delta U^l(d)$ for all $d \geq 0$.*

Proof. Part (a) is straightforwardly verified by differentiating $\Delta U^h(d)$ and $\Delta U^l(d)$ with respect to d. Part (b) in turn follows from a direct comparison of the expressions for $\Delta U^h(d)$ and $\Delta U^l(d)$. ■

The functions $\Delta U^h(d)$ and $\Delta U^l(d)$ and are shown in Figure 2. The reason why these functions are increasing functions of the mass of dealers d is simply that with a greater mass of dealers there is a greater likelihood $m(d, a^*)$ for an entrepreneur with a good asset to be matched with an informed dealer. Thus, an entrepreneur deviating from a low-origination equilibrium a_l by choosing a_h is more likely to get rewarded with a match in the OTC market in the event that he has a good asset. Therefore his incremental payoff from deviating is larger. As for an entrepreneur deviating from a high-origination equilibrium a_h by choosing a_l, the higher is d the more good assets get skimmed in the OTC market, which results in a lower price p in the organized market at which the entrepreneur can sell his bad asset. This is why $\Delta U^h(d)$ is also increasing in d. Finally, intuition suggests that the effect of an increase in d on the price in the organized market p is likely to be smaller than the effect on the matching probability $m(d, a^*)$ in the OTC market. This is confirmed by the lemma and is why $\Delta U^h(d)$ $< \Delta U^l(d)$ for all $d \geq 0$.

Next, if we define $\widehat{d^h}$ and $\widehat{d^l}$ respectively by the following equations

$$\Delta U^h(\widehat{d^h}) = \psi \qquad \text{and} \qquad \Delta U^l(\widehat{d^l}) = \psi,$$

we are able to establish our first major characterization of equilibrium occupational choice in period 0 in the following proposition (see also Figure 3).

Proposition 5 *(a)* $\widehat{d^l} < \widehat{d^h}$*. (b) A low-origination-effort equilibrium can only be supported for* $d \in [0, \widehat{d^l}]$ *and in particular no low effort equilibrium exists when* $\psi < (1-\pi)\rho\Delta a\,(\gamma - 1)$*. (c) A high-origination-effort equilibrium can only be supported for* $d \in [\widehat{d^h}, 1]$*; and (d) there is no equilibrium with* $d \in (\widehat{d^l}, \widehat{d^h})$*.*

Proof. Part (a) follows directly by comparing the solutions to $\Delta U^h(d) = \psi$ and $\Delta U^l(d) = \psi$. Part (b) follows from the fact that for any $d > \widehat{d^l}$ the incentive constraint (12) is not satisfied. Similarly, part (c) follows from the fact that for any $d < \widehat{d^h}$ the incentive constraint (9) is not satisfied. Finally, since neither incentive constraints (12) or (9) are satisfied for $d \in (\widehat{d^l}, \widehat{d^h})$, part (d) immediately follows. ■

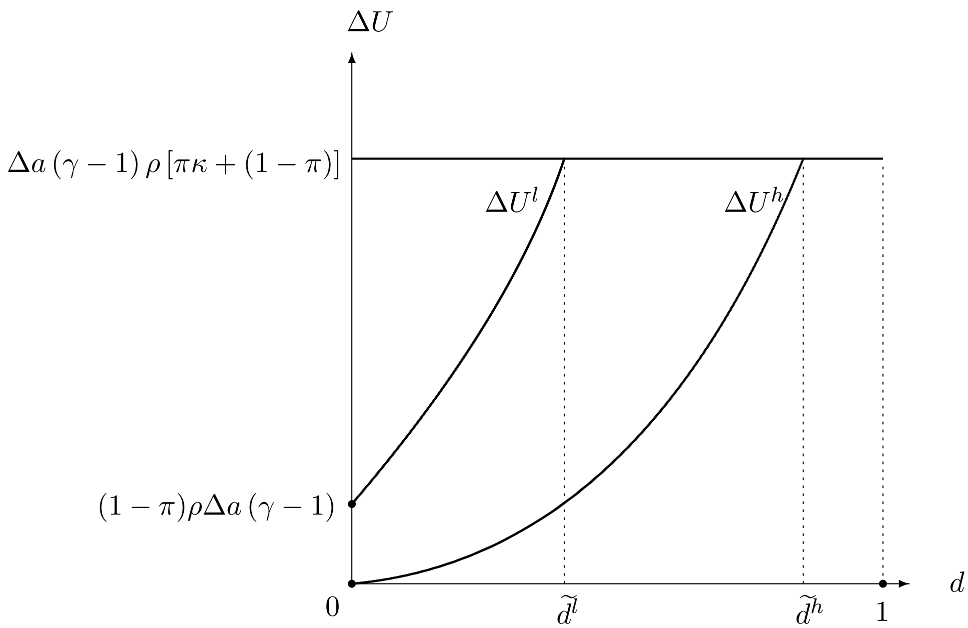

Figure 2: **Incentive compatibility for the low and high effort profile.** The functions $\Delta U^l\,(d)$ and $\Delta U^h\,(d)$ are increasing in d and $\Delta U^l\,(d) > \Delta U^h\,(d)$. $\widetilde{d}^l$ and $\widetilde{d}^h$ are the first measures of dealers for which $m\left(a_l, \widetilde{d}^l\right) = 1$ and $m\left(a_h, \widetilde{d}^h\right) = 1$, respectively.

Another complicating factor for the existence of either type of equilibrium is that the expected utility of any dealer d in period 0, $V(d \mid a^*, d^*)$, is increasing in the equilibrium measure of dealers d^* in the market. This property of dealer payoffs can give rise to multiple equilibria of either type. Consider first the low-origination-effort equilibrium type (a_l, d^l) such that $d^l \in [0, \widehat{d}^l]$. If all type 2 agents expect no dealers to enter the OTC market in period 0 this may well be a self-fulfilling equilibrium outcome if $V(0 \mid a_l, 0) \leq U(a_l \mid a_l, 0)$. But this low-origination-effort equilibrium with no OTC markets can also co-exist with potentially two other low-origination-effort equilibria with a positive mass of dealers d_1^l and $d_2^l > d_1^l$, where d_1^l is an *unstable equilibrium* such that $V(d_1^l \mid a_l, d_1^l) = U(a_l \mid a_l, d_1^l)$ and $V(d \mid a_l, d) \geq U(a_l \mid a_l, d)$ for all $d \in (d_1^l, d_2^l)$, and d_2^l is a *stable equilibrium* such that $V(d_2^l \mid a_l, d_2^l) = U(a_l \mid a_l, d_2^l)$ and $V(d \mid a_l, d) < U(a_l \mid a_l, d)$ for $d > d_2^l$. This is indeed what we find, as the following numerical example illustrates.

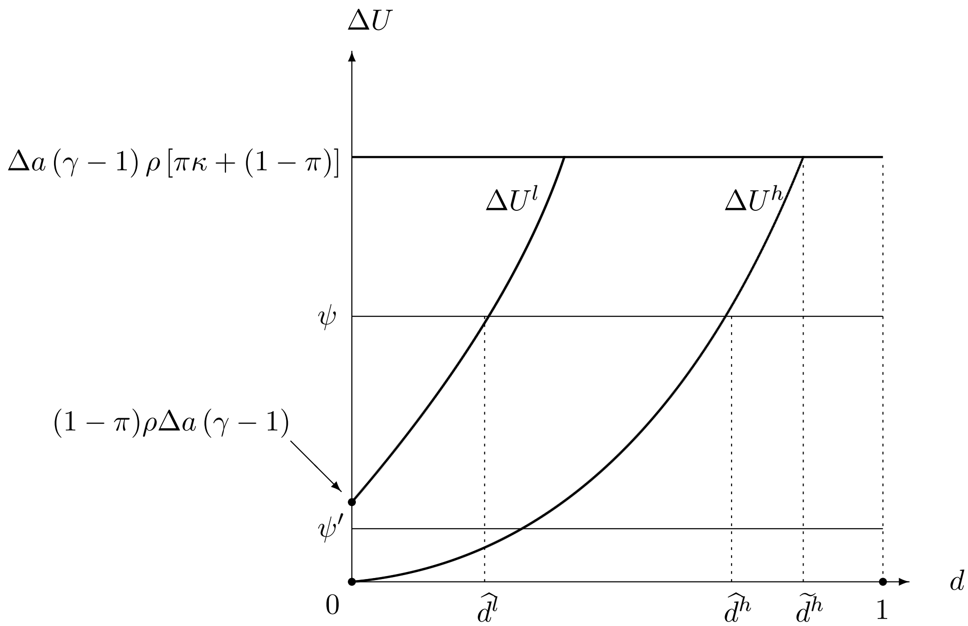

Figure 3: **Incentive compatibility.** Incentive compatibility for the low and high effort profile. When the cost of providing the high effort is given by ψ then a candidate low effort equilibrium is only incentive compatible if and only of $d \in [0, \widehat{d}^l]$ and a candidate high effort equilibrium is only incentive compatible if and only if $d \in [\widehat{d}^h, \widetilde{d}^h]$

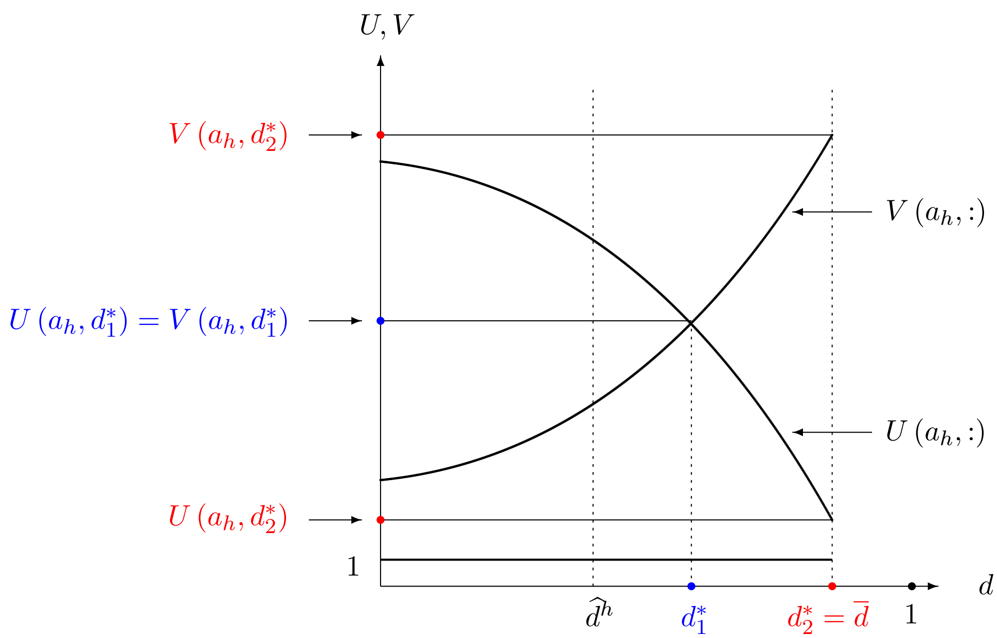

Figure 4: **Example 1 - High effort equilibria.** When the information cost function is $\varphi(j) = 0$ for $j \leq \overline{d}$ and $\varphi(j) = +\infty$ for $j > \overline{d}$ there might two possible high effort equilibria. d_1^* is the measure of dealers in the unstable high effort equilibrium and $d_2^* = \overline{d}$ in the stable high effort equilibrium.

Example 1 (High origination-effort equilibria). Consider the following parameter values

$$a_h = .75 \quad a_l = .5 \quad \gamma = 1.5 \quad \rho = .8 \quad \kappa = .25 \quad \psi = .0013 \quad \pi = .5 \tag{13}$$

We also assume

$$\varphi(d) = 0 \quad \text{for} \quad d \in [0, .35] \quad \text{and} \quad \varphi(d) = +\infty \quad \text{for} \quad d > .35 \tag{14}$$

In this case

$$\widetilde{d}_h = .4286 \qquad \text{and} \qquad \widehat{d}_h = .0536 \tag{15}$$

Incentive compatibility region for high origination-effort requires that $d \geq \widehat{d}^h = .0536$, as shown in Figure 4. It can be shown that no low effort allocation is incentive compatible as $(1-\pi)\rho\Delta a\,(\gamma - 1) > \psi = .0013$ (see Figure 3), which implies $\Delta U^l(d) > \psi$ for all $d \geq 0$.

As for high origination-effort equilibria, there are two and they are shown in Figure 4. First, there is an unstable equilibrium with $d_1^h = .3101$ in which the agents $j \leq \overline{d}$ are indifferent between becoming entrepreneurs or dealers. Second, there is a stable equilibrium with $d_2^h = .35$, in which dealers are *strictly* better off as such than as entrepreneurs. Notice that all agents who can become dealers are dealers in equilibrium and thus our economy is at a corner.

Note that the price of assets in the OTC market in the unstable equilibrium is then

$$p^d\left(a_h, d^h = .3\right) = 1.0184, \tag{16}$$

so that a dealer needs some leverage in order to finance the purchase of the asset. In contrast, in the stable equilibrium leverage is not needed as

$$p^d\left(a_h, d^h = .35\right) = .9833. \tag{17}$$

which is less than their endowment. $\diamondsuit$

We turn next that to the issue of the low effort equilibrium.

Example 2 (The low effort equilibrium). Consider the following example

$$a_h = .75 \quad a_l = .55 \quad \gamma = 1.5 \quad \rho = .8 \quad \kappa = .25 \quad \psi = .0475 \quad \pi = .5 \tag{18}$$

We also assume

$$\varphi(d) = 0 \quad \text{for} \quad d \in [0, .15] \quad \text{and} \quad \varphi(d) = +\infty \quad \text{for} \quad d > .15 \tag{19}$$

It can be numerically shown that all possible occupational choices are incentive compatible in that

$$\Delta U^l(d) < \psi \qquad \text{for all} \qquad d \in [0, .15]. \tag{20}$$

As shown in Figure 5, there are then three (low origination-effort) equilibria. First there is an efficient stable equilibrium where $d_l^* = 0$. Indeed, notice that when there are no dealers $U(a_l \mid a_l, 0) > V(0 \mid a_l, 0)$. Second, there is an inefficient unstable equilibrium with a measure of informed dealers $d_1^l = .0781$ with $U(a_l \mid a_l, .0781) = V(.0781 \mid a_l, .0781)$ and, finally, there is an inefficient stable equilibrium with $d_2^l = .15$ where $U(a_l \mid a_l, .15) < V(.15 \mid a_l, .15)$. $\diamondsuit$

5 Welfare: the inefficiently large size of the financial sector

The multiplicity of equilibria highlighted in the previous section reveals the possibility that two economies with the same underlying characteristics may have financial sectors of very different sizes. We have also shown that there is a fundamental complementarity in our model between the real and the financial sectors. An economy which generates a lot of high quality assets ($a^* = a_h$) is also an economy with a large financial sector ($d \geq \widehat{d}^h$), and an economy which originates only a few high quality assets ($a^* = a_l$) is one with a relatively small financial sector ($d \leq \widehat{d}^l$). One would be tempted to conclude from these observations that our economy tends to generate an efficient occupational choice between the real and financial sectors.

Our main result of the paper, however, is to show that our equilibria (with one exception) are all constrained inefficient in the sense that they tend to have an excessively large financial

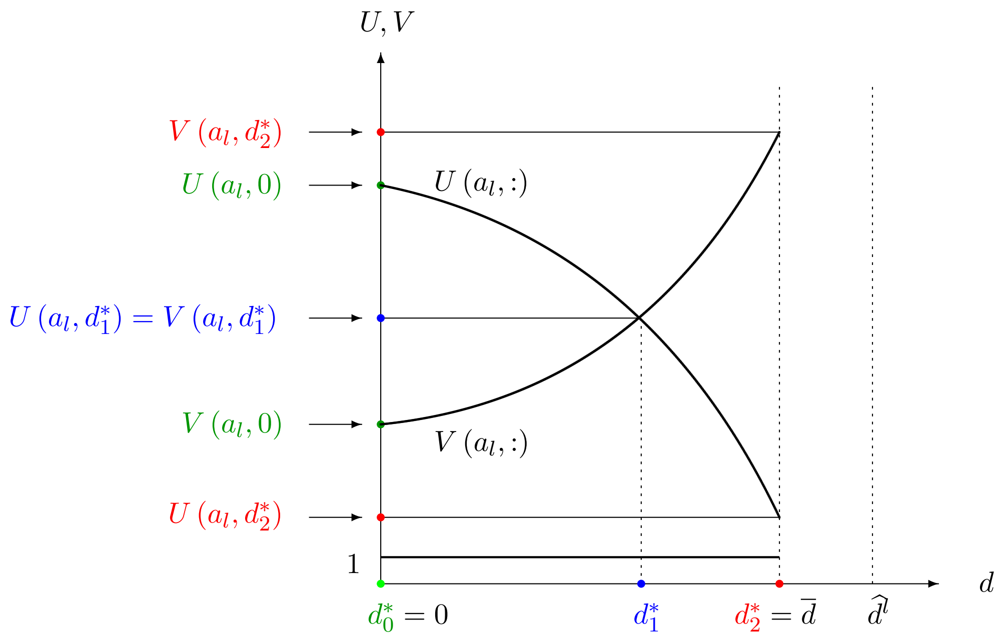

Figure 5: **Low effort equilibria.** When the information cost function is $\varphi(j) = 0$ for $j \leq \overline{d}$ and $\varphi(j) = +\infty$ for $j > \overline{d}$ there might three possible high effort equilibria, one unstable and two stable equilibria. $d_0^* = 0$ obtains in a low effort equilibrium with no dealers. d_1^* is the measure of dealers in the unstable low effort equilibrium and $d_2^* = \overline{d}$ in the stable low effort equilibrium, with a strictly positive measure of dealers.

sector. More precisely, our equilibria have inefficiently large OTC markets as we show in the proposition below.

Our notion of *constrained efficiency* is based on the standard idea that the social planner should not have an informational advantage relative to an uninformed market participant. Thus, we only allow the planner to dictate the occupation of type 2 agents and we do not let the planner make any decisions based on the information obtained by informed dealers. The planner's problem in period 0 is then to pick the measure d of type 2 agents that maximizes ex-ante social surplus. There are only two possible socially efficient allocations of type 2 agents in period 0, $d^* \in \{d^{l*}, d^{h*}\}$. If it is socially efficient to implement the low origination-effort a_l, then the efficient allocation consistent with that outcome is given by $d^{l*} \in [0, \hat{d}^l]$, and if it is socially efficient to implement the high origination-effort a_h, then the socially efficient allocation is $d^{h*} \in [\hat{d}^h, \tilde{d}^h]$.

It is straightforward to verify that for the parameter values given in **Example 1** above the socially efficient origination effort is a_h:

$$[\rho(1+a_h(\gamma-1))-\psi]\left(1-\widehat{d}_h\right)-\varphi\widehat{d}_h-\rho(1+a_l(\gamma-1))=.0399 \quad (21)$$

Similarly, for the parameter values given in **Example 2** the socially efficient origination effort is a_l:

$$[\rho(1+a_h(\gamma-1))-\psi]\left(1-d_h^{fb}\right)-\varphi d_h^{fb}-\rho(1+a_l(\gamma-1))=-.44. \quad (22)$$

Obviously then, the socially efficient allocation in example 1 is $d^{l*}=0$, and the efficient allocation in example 2 is $d^{h*}=\hat{d}^h$. That is, the socially efficient allocation $d^*\in\{d^{l*},d^{h*}\}$ is given by the smallest measure of informed dealers required to support entrepreneur incentives towards respectively a low and high origination-effort in the real sector.

Comparing constrained efficient and equilibrium allocations, the next set of propositions establish the main result of our analysis, namely that in equilibrium the OTC market is inefficiently large.

Proposition 6 *(a) All high origination-effort equilibria are generically constrained inefficient in that $d^{h*}=\hat{d}^h$, (b) All low origination-effort equilibria with a strictly positive measure of dealers are also constrained inefficient.*

Proof. Part (a) immediately follows from the observation that all high origination-effort equilibria have a measure of informed dealers $d\in[\hat{d}^h,\tilde{d}^h]$. It would then be pure coincidence if the solution $d_1^h\in[\hat{d}^h,\tilde{d}^h]$ (when it exists) to the equation $V(d_1^h\mid a_h,d_1^h)=U(a_h\mid a_h,d_1^h)$ is such that $d_1^h=\hat{d}^h$. Part (b) immediately follows from the observation that $d^{l*}=0$. ■

Note that when entrepreneurs' effort cost ψ is low enough then, as Figure 2 illustrates, only high origination-effort equilibria can exist, in which case all equilibria have the property that they generate an excessively large financial sector. The basic economic logic underlying this result is that *at the margin* the ex-ante payoff from becoming a dealer is increasing in the size of the OTC market. Therefore, if entry into the OTC market is left unchecked there will tend to be too much entry.

It is not just that a given equilibrium will tend to have an excessively large OTC market, but also equilibria can be ranked in terms of ex-ante social efficiency. That is, the equilibria with the larger OTC markets are also the most inefficient ones (conditional on the same origination effort).

Proposition 7 *Equilibria with the same origination-effort can be ranked by total ex-ante social surplus in decreasing order of the equilibrium size of the OTC market.*

Proof. Given our assumptions on $\varphi(d)$, there can be up to three low origination-effort, and up to two high origination-effort equilibria. As the efficient allocation under low origination-effort is $d^{l*} = 0$, as $d_1^h \geq \hat{d}^h$, and as total ex-ante social surplus is strictly decreasing in d, the proposition immediately follows from these observations. ■

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

6 Appendix

To be completed.

1. Necessary condition guaranteeing that $p^d - 1 \leq \rho$, or

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

Substituting for p in in equation (5) this inequality becomes

$$\kappa\gamma\rho + (1-\kappa)\frac{\rho[a(1-m)\gamma + (1-a)]}{1-am} \leq 1 + \rho$$

and substituting for m in equation (4) we obtain a restriction on parameter values which we impose for the remainder of our analysis, for $a \in \{a_l, a_h\}$:

$$\textbf{Assumption A}: \kappa\gamma + (1-\kappa)\frac{[\gamma[a(1-d)\pi - d(1-\pi)] + (1-d)\pi(1-a)]}{\pi - d} \geq 1 + \frac{1}{\rho}.$$

(Although d is an endogenous variable we can ensure that d remains relatively small in equilibrium by assuming that $\overline{d}$ is small).