Financial Relationships and the Limits to Arbitrage

Jiro E. Kondo*

Dimitris Papanikolaou†

May 2012

Abstract

We propose a new foundation for the limits to arbitrage based on *financial relationships* between arbitrageurs and banks. Financially constrained arbitrageurs may choose to seek additional financing from banks who can understand their strategies. However, a hold-up problem arises because banks cannot commit to provide capital and have the financial technology to profit from the strategies themselves. Wary of this, arbitrageurs will choose to stay constrained and limit their correction of mispricing unless banks have sufficient reputational capital. Using the framework of stochastic repeated games, we show that this form of limited arbitrage arises when mispricing is largest and becomes more substantial as the degree of competition between banks intensifies and arbitrageur wealth increases.

*McGill University

†Kellogg School of Management and NBER

1 Introduction

Arbitrageurs rely on investment banks for a substantial portion of their financing. However, when accessing this source of finance, they run the risk of being expropriated through activities such as front-running. So why do they use this source of funding instead of other forms of external finance? First, we argue that investment banks face less severe adverse selection problems. Namely, they are a type of *informable* financier because arbitrageurs can reveal their private information to them. Second, we show that a bank's reputation can allow it to commit to not exploiting arbitrageurs. However, this commitment is limited. This paper studies the role of financial relationships in enabling informable finance and, in doing so, proposes a new foundation for the limits to arbitrage.

Using a stochastic repeated game that arises from time-variation in mispricing, we develop further predictions about when arbitrage becomes limited. Namely, we show that limited arbitrage occurs when initial mispricing is most severe. This observation cannot be obtained from standard repeated game models and is robust to realistic assumptions about contractibility. We also demonstrate that our limits to arbitrage problem becomes more substantial as competition increases between investment banks in attracting business from arbitrageurs and, surprisingly, as arbitrageur wealth increases. This last finding can partially explain why arbitrageurs occasionally choose to refuse additional capital from investors.

Existing models of limited arbitrage typically introduce a source of non-fundamental risk and conclude that arbitrage opportunities are risky investments (e.g., De Long et al. (1990) and Shleifer-Vishny (1997)). In these settings, arbitrageurs are reluctant to trade aggressively against mispricing out of fear that it will worsen and lead them to liquidate their positions at a loss. This prediction is driven by two assumptions: arbitrageurs are subject to financial constraints and potential capital providers cannot understand the arbitrageur's investment opportunities. The latter assumption is called the "separation of brains and capital."

Yet, the view that *all* potential financiers cannot understand the arbitrageurs' strategies is extreme. Though arbitrageurs have significantly more expertise than the typical investor, some knowledgeable providers of capital have the experience and skill required to adequately evaluate their opportunities. For instance, investment banks gain similar expertise through proprietary trading while also funding arbitrageurs through prime brokerage operations. In the presence of such informable finance, one might think that arbitrage would not be limited. Regardless of recent performance, arbitrageurs could exploit new opportunities by simply revealing their strategies to banks and then borrowing funds or getting them to reduce margin requirements.

However, this logic is flawed. Since courts, like most investors, cannot properly evaluate the communication between arbitrageurs and banks, contracts are incomplete and banks cannot explicitly commit through formal contracts to provide any capital ex-post. Moreover, the experience that allows banks to understand the arbitrageurs also gives them access to similar, if not superior, financial technology to execute these strategies. This creates a hold-up problem: after the arbitrageur reveals his information to the bank, what prevents the latter from providing no capital and undertaking the profitable transactions for herself?[1]

In a one-shot transaction, the answer is nothing. Banks do not lose anything when holding up arbitrageurs and therefore cannot credibly commit to making arbitrageurs better off when they reveal their strategies. Arbitrageurs understand this and refuse to share information with them. Given this decision, their only resort is to seek financing from the uninformed. The separation of brains and capital arises endogenously.

However, when similar interactions occur repeatedly, a bank's concern for its reputation can persuade the arbitrageur to reveal his information. We consider a sequence of arbitrageurs, each with knowledge of a different arbitrage opportunity, who interact with infinitely lived banks. In equilibrium, arbitrageurs and banks optimally collude whenever the latter can commit to acceptable behavior. Specifically, arbitrageurs reveal their information and allow banks to keep some profits for themselves while trading is structured so that total surplus is maximized.

Nevertheless, the value of financial relationships as a disciplinary mechanism is limited: it's value is proportional to the average profitability of an arbitrage opportunity whereas a bank's profits from fully expropriating the arbitrageur varies with the current level of mispricing. If the most profitable arbitrage opportunities are sufficiently superior to the average one, banks won't always be able to commit to cooperation. This problem occurs exactly when mispricing is largest and communication would create the largest surplus between the two parties.

In addition to the limits to arbitrage literature, our paper is also related to several other strands of finance. First, it provides an additional example where some investors can take advantage of others through knowledge of their proprietary strategies and trading needs. Brunnermeier-Pedersen (2005) study predatory trading in response to the predictable activity of distressed investors. Ko (2009) considers front-running by banks as a cost to arbitrageurs when disclosing their risk profiles and shows that this can lead to endogenous concentrations of risk. His analysis differs from ours because it is static and does not focus on financial relationship or the limits to arbitrage. Finally,

[1]Relatedly, information provided by the arbitrageur to the bank's prime brokerage division can be leaked to the bank's proprietary traders.

our work is related to research on the efficiency benefits of bank reputation (e.g., Sharpe (1990), and Chemmanur-Fulghieri (1994a, 1994b)).[2]

This paper is also closely related to the general setting of financing innovation and selling ideas when intellectual property rights are imperfect. Because of informational asymmetries, potential financiers and buyers are unlikely to offer a fair price for valuable innovations and good ideas unless details are provided to them ex-ante. However, once they have this knowledge, they may effectively own all its productive use and have little incentive to pay for it ex-post. This hold-up problem, which is identical to ours, is known as the fundamental paradox (Arrow (Ch. 6, 1971)).

There is a large literature that explores ways of mitigating this problem. Anton-Yao (1994) show that the existence of competition among potential buyers can improve efficiency when the entrepreneur can threaten to reveal his idea to a second buyer when the first refuses to pay him appropriately. Rajan-Zingales (2001) study how organizational hierarchy can be used to minimize the problem of information leakage. Rather than looking at commitment, Anton-Yao (2002, 2004) explore the use of partially expropriable disclosures to signal project value. Nevertheless, none of these models achieve the first-best. In independent work, Hellman-Perotti (2011) show that firm reputation can foster more efficient innovation. A reinterpretation of our model adds to their observation by remarking that the reputation mechanism is limited in a particular way: it fails to achieve first-best implementation for the most valuable ideas.

This paper has two main contributions. First, it provides new economic foundations for the limits to arbitrage: fear of opportunism by informable financiers. This complements existing theory by identifying conditions under which the separation of brains and capital would endogenously arise. The second contribution is broader. Namely, this paper is one of the first to explicitly incorporate financial relationships in the asset pricing and microstructure literature. Other work includes Benabou-Laroque (1992), Desgranges-Foucault (2005), Bernhardt et al. (2005), and Lobo et al. (2007). The observation that trade and financial interaction among agents, especially large institutions, is not anonymous is obvious, yet its effect on pricing has remained largely unexplored. It would be interesting to investigate how this aspect of trade ameliorates or worsens informational asymmetries and incentive problems in financial markets. Such analysis could also lead to a better understanding of the existing institutional structure of the securities industry. We open a discussion on the market structure of arbitrage later in this paper by commenting on the role that certain arbitrage institutions, like fund-of-funds and

[2]Our model is also similar to Rotemberg-Saloner (1986) who study a stochastically repeated game with i.i.d. variation. However, all the strategic players in their setting are homogeneous while we assume heterogeneity between the arbitrageur and the bank in the form of initial information asymmetry.

seeders, might play in overcoming the hold-up problem in our model.

The remainder of this paper proceeds as follows. Section 2 presents the model and its assumptions. Section 3 derives a characterization of the equilibria and their general properties. Section 4 considers a simple refinement based on the degree of banking competition and presents examples of its implication on limited arbitrage. Additional extensions and discussion are in section 5. All proofs are gathered in the Appendix.

2 The Model

We consider a parsimonious framework where arbitrage opportunities are riskless and converge immediately following a round of trading. An arbitrage opportunity is modeled as an extensive form game where initial mispricing is generated at $t = 0$, relational interactions between the arbitrageur and the investment bank (bank) occur at $t = 1$, one round of trading takes place at $t = 2$, and terminal payoffs are realized at $t = 3$. In this section, all variables, including trading profits, are observable but not verifiable. We relax this assumption in section 5.1 and show that our results remain largely unaffected by contracting if we allow for adverse selection. We study both the one-shot and infinitely repeated versions of this game. The stage game in the latter version is called an arbitrage opportunity cycle.

There are N risky assets in the economy and a riskless asset with a rate of return normalized to 0. All agents have a common prior over the terminal payoffs of these assets, namely that they are imperfectly correlated with identical means, $\overline{v}$. The arbitrageur receives a private signal at $t = 0$, informing him that two assets, A_1 and A_2, have identical terminal payoffs. We will focus on equilibrium in the markets for these two assets. For simplicity, we assume that the universe of assets is sufficiently large that all other investors cannot infer the arbitrageur's information from prices as the posterior probability that a given asset is part of the arbitrageur's signal is always negligible.

In each market, there are two types of non-strategic investors who place demands: noise traders and long-term traders. Noise traders buy and sell randomly in each market and are responsible for the existence of arbitrage opportunities. Their trading in A_1 and A_2 is X_{N1} and X_{N2}, respectively. We denote by F the cumulative distribution of the noise trader spread, $\Delta X_N \equiv |X_{N1} - X_{N2}|$, and assume that it has a finite second moment. In this environment, ΔX_N can be seen as a measure of initial mispricing. Long-term traders submit downward sloping linear demands for each risky asset,

$$X_{LRi} = \frac{1}{\lambda}\Big(\overline{v} - p_i\Big), \tag{1}$$

where $\lambda > 0$. This specification of residual demand curves is standard in the literature (e.g., Brunnermeier-Pedersen (2004) and Xiong (2001)). Various interpretations can be given to the long-term traders. For example, they can be viewed as risk averse market makers (e.g., Grossman-Miller (1988)) or uninformed investors fearing exploitation by informed ones (e.g., Grossman-Stiglitz (1980) and Kyle (1985)).

The main analysis focuses on the behavior of a sequence of short-lived arbitrageurs and a long-lived bank. Both groups are risk-neutral and strategic. The bank lives forever and has a discount factor δ.[3] Arbitrageurs live for one cycle and place convergence trades on A_1 and A_2.[4] They have limited wealth and face a default financial constraint of the form,

$$X_A \leq M_L, \tag{2}$$

where X_A is the amount of convergence trading undertaken by the arbitrageur. This constraint captures the arbitrageurs' limited access to initial funding from both investors and banks, the latter through margin financing. Microeconomic foundations for this type of constraint are well established (e.g., Kiyotaki-Moore (1997) and Gromb-Vayanos (2002)).

In each arbitrage opportunity cycle, the arbitrageur observes ΔX_N and can trade without revealing his strategy to the bank. This choice is denoted by $R = 0$. The bank does not trade in this case and prices are determined by market clearing:

$$p_1 = \overline{v} + \lambda(X_{N1} - X_A) \quad \text{and} \quad p_2 = \overline{v} + \lambda(X_{N2} + X_A). \tag{3}$$

Alternatively, the arbitrageur can choose to share his information to the bank in the hope of negotiating an increase in his trading capacity. This increase can be achieved with an infusion of capital or, as is more common, a renegotiation of terms in the margin agreement. This choice is denoted by $R = 1$. Unfortunately, the bank cannot commit ex-ante to alleviate the arbitrageur's financial constraint and may select any position limit satisfying $M \geq M_L$. It can also chose to front-run the arbitrageur by trading an amount $X_B \geq 0$ ahead of him. The arbitrageur observes X_B prior to choosing his own position, $X_A \leq M$, and all trades clear simultaneously at the market clearing price:

$$p_1 = \overline{v} + \lambda(X_{N1} - X_A - X_B) \quad \text{and} \quad p_2 = \overline{v} + \lambda(X_{N2} + X_A + X_B).^5 \tag{4}$$

[3]Beyond bank impatience, the discount factor can also serve as a proxy for the frequency of discovering new arbitrage opportunities or as a reduced form for relevant elements that are not explicitly modelled, like bank risk aversion or agency problems within the bank.

[4]A convergence trade is defined as a long position in A_i and an equally short position in A_j. In a different setting, Xiong (2001) also studies equilibrium mispricing under the same assumption that arbitrageurs are restricted to such trades.

[5]Neither the observability of X_B nor the simultaneity of trades are critical to our qualitative results.

Profits for the arbitrageur and bank are given by $\Delta p \cdot X_A$ and $\Delta p \cdot X_B$, respectively, where $\Delta p \equiv p_1 - p_2$. See Figure 1 for an illustration of this stage game:

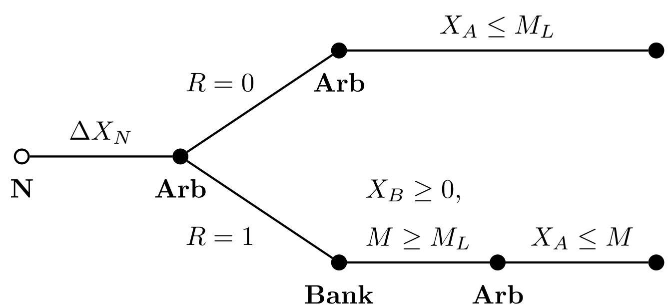

Figure 1: An arbitrage opportunity cycle.

Throughout the statement of our results, we will denote the equilibrium choice of any variable X by X^*.

3 Equilibrium

3.1 Benchmark Equilibria

Prior to determining the equilibria of the game, it is instructive to look at two benchmark outcomes. The first has no arbitrageurs or banks and is referred to as the no arbitrageur equilibrium (NA). The second case assumes that arbitrageurs have unlimited wealth and face no financial constraints or, equivalently, that there are no agency problems between the arbitrageurs and the bank. This benchmark is called the first-best equilibrium (FB).

Proposition 1: *In the no arbitrageur equilibrium, the price spread is given by:*

$$\Delta p^{NA} = \lambda \Delta X_N. \tag{5}$$

Meanwhile, in the first-best equilibrium, the arbitrageur's demand is given by:

$$X_A^{FB} = \frac{1}{4} \Delta X_N, \tag{6}$$

while the resulting price spread and profit are equal to:

$$\Delta p^{FB} = \frac{\lambda}{2} \Delta X_N \qquad \text{and} \qquad \Pi^{FB} = \frac{\lambda}{8} \Delta X_N^2. \tag{7}$$

In the absence of arbitrageurs, the price spread arises because long-term traders absorb ΔX_N more units of one asset than the other and require an additional premium of λ per unit of demand imbalance. In the first-best case, since the arbitrageur acts strategically, he eliminates only half of the relative mispricing.

3.2 One-Shot Game

We consider subgame perfect equilibria of the one-shot game.

Proposition 2: *There is a unique SPE and the arbitrageur's equilibrium strategy is:*

$$R^* = 0 \qquad \text{and} \qquad X_A^* = \begin{cases} \frac{1}{4}\Delta X_N & \text{if } R = 0 \text{ and } \Delta X_N \leq 4M_L \\ \frac{1}{4}\Delta X_N - \frac{1}{2}X_B & \text{if } R = 1 \text{ and } X_B \geq \frac{1}{2}\Delta X_N - 2M_L \\ M_L & \text{otherwise} \end{cases}, \tag{8}$$

while the bank's equilibrium strategy is:

$$M^* = M_L \qquad \text{and} \qquad X_B^* = \begin{cases} 0 & \text{if } R = 0 \\ \frac{1}{4}\Delta X_N & \text{if } R = 1 \text{ and } \Delta X_N \leq (4+2\sqrt{2})M_L \\ \frac{1}{4}\Delta X_N - \frac{1}{2}M_L & \text{if } R = 1 \text{ and } \Delta X_N \geq (4+2\sqrt{2})M_L \end{cases}. \tag{9}$$

Since the bank cannot commit to relax the arbitrageur's financial constraint or to not copy his strategy, it always sets $M = M_L$ and $X_B > 0$ following communication. Because the bank has price impact, this unambiguously makes the arbitrageur worse off and he prefers not to reveal his information. Arbitrage activity is limited. This also provides more precise foundations for Shleifer-Vishny's limited arbitrage by identifying an explanation for the separation of brains and capital. In equilibrium, both parties are worse off. By committing to loosen the arbitrageur's financial constraint and limit its front-running, the bank could achieve positive profits while still making the arbitrageur better off.

Prior to analyzing the repeated game, it is useful to determine the profits to the arbitrageur and the bank in the continuation games from the static case following communication and no communication. The case following communication will determine the bank's profits when optimally deviating from the relational contract,

$$\Pi_B^d = \begin{cases} \frac{\lambda}{16}\Delta X_N^2 & \text{if } \Delta X_N \leq (4+2\sqrt{2})M_L \\ \\ \frac{\lambda}{8}\Delta X_N^2 - \frac{\lambda}{2}M_L\Delta X_N + \frac{\lambda}{2}M_L^2 & \text{if } \Delta X_N \geq (4+2\sqrt{2})M_L \end{cases}. \tag{10}$$

The case without communication will provide the arbitrageur's outside option:

$$\overline{\Pi}_A = \begin{cases} \frac{\lambda}{8}\Delta X_N^2 & \text{if } \Delta X_N \leq 4M_L \\ \\ \lambda M_L \Delta X_N - 2\lambda M_L^2 & \text{if } \Delta X_N \geq 4M_L \end{cases}. \tag{11}$$

For details, see Lemmas 3 and 4 in the Appendix.

3.3 Repeated Game

In the infinitely repeated version of the game, we need to make additional assumptions on the information that each arbitrageur has about the past behavior of the bank. Specifically, arbitrageurs know the full history of the bank's behavior. This degree of knowledge could be rationalized if $\Delta X_{N,t}$ and $\Pi_{A,t}$ became known to the market at $t+1$, perhaps because arbitrages only remain private for a limited time and the arbitrageur's terminal payoff is public information. We could also extend the model to incorporate imperfectly observable actions in line with Abreu-Pearce-Stacchetti (1990) and Fudenberg-Levine-Maskin (1994). Nevertheless, this would only amplify our limits to arbitrage problem because the bank would occasionally get away with misbehavior and receive blame when it behaves.

We also assume that $\Delta X_{N,t}$ is independent and identically distributed over time. The independence assumption is reasonable because any predictable component would be known to the entire market and eliminated through competition. The arbitrageur can only expect to profit on the surprise component of $\Delta X_{N,t}$.

Relational contracts are promises the bank makes to the arbitrageur regarding her behavior. However, these promises cannot be enforced by a court. In a given cycle, promises take the form of functions $M_t^c : \Re \rightarrow [M_L, \infty)$ and $X_{B,t}^c : \Re \rightarrow [0, \infty)$ that specify the bank's actions as a function of $\Delta X_{N,t}$. To conserve on notation, we will suppress the dependence of M_t^c and $X_{B,t}^c$ on $\Delta X_{N,t}$. Arbitrageurs cannot make any credible promises to the bank because they only live for one cycle.

We consider efficient subgame perfect equilibria of the repeated game.

Definition (Efficient Equilibria): *An SPE of the repeated game with payoffs* $(\Pi_{A,t}^1, \Pi_{B,t}^1)$ *is efficient if and only if there does not exist another SPE of the game with payoffs* $(\Pi_{A,t}^2, \Pi_{B,t}^2)$ *such that: (i) for every* $\Delta X_{N,t}$, $\Pi_{A,t}^2 \geq \Pi_{A,t}^1$, *and (ii)* $V_t^2 \geq V_t^1$ *where* V_t *is the value of the relationship to the bank at* t*:*

$$V_t = E_t \left[\sum_{j=1}^{\infty} \delta^{t+j} \Pi_{B,t+j} \right]. \tag{12}$$

We discard SPEs that are not efficient because the bank and the arbitrageur can agree to alter their component of the relational contract immediately after $\Delta X_{N,t}$ is realized.[6] Such an agreement is possible if the arbitrageur can be made weakly better off, regardless of initial mispricing, while also improving the bank's continuation payoff. This is essentially an interim Pareto-optimality criterion (see Brunnermeier (2001)).

In equilibrium, the arbitrageur will only choose $R_t^* = 1$ if the bank can credibly commit to satisfy his individual rationality constraint:

$$\Pi_{A,t}(R_t = 1) \geq \overline{\Pi}_{A,t}. \tag{13}$$

Similarly, the bank's choice of $(M_t^*, X_{B,t}^*)$ can be restricted to those specified by the relational contract, $(M_t^c, X_{B,t}^c)$, and the optimal deviation levels, $(M_t^d, X_{B,t}^d)$. She chooses to cooperate if and only if:

$$\Pi_{B,t}^c + V_t^* \geq \Pi_{B,t}^d \tag{14}$$

where V_t^* is the continuation value of the financial relationship to the bank in the particular SPE at t. Due to the i.i.d. and finite second moment assumptions, this value is bounded. Implicit in the bank's incentive compatibility constraint is the assumption of maximal punishment by arbitrageurs. Such punishment is credible if there are other equally qualified banks in the market and there is no cost to moving the relationship from one bank to another.

Efficient SPEs induce a structure of optimal collusion between the arbitrageur and the bank. The first element of optimal collusion is illustrated in Lemma 5:

[6]Presumably, this would not be considered a deviation by future generations of arbitrageurs since it does not reflect an action that caused a welfare loss to a previous arbitrageur (i.e. it only produces mutually beneficial gains).

Lemma 5: *If $R_t^* = 1$, then an efficient equilibrium satisfies:*

$$X_{B,t}^* + M_t^* = \frac{1}{4}\Delta X_{N,t} \tag{15}$$

and

$$X_{A,t}^* = M_t^* \tag{16}$$

That is, the arbitrageur and the bank optimally collude to achieve first-best total profits.

When communication occurs, the total demand of the arbitrageur and bank equals that of the first-best equilibrium from Proposition 1. As a result, price spreads are as in the first-best and profits shared between the arbitrageur and bank are maximized. Put simply, if the arbitrageur and the bank are going to collude, they will do so effectively.

Plugging $X_{A,t}^*$ from the lemma into the arbitrageur's profit function yields:

$$\Pi_{A,t}(R_t = 1) = \frac{\lambda}{2}\Delta X_{N,t}\left(\frac{1}{4}\Delta X_{N,t} - X_{B,t}^*\right). \tag{17}$$

Along with the arbitrageur's outside option, $\overline{\Pi}_{A,t}$, we get an upper bound on the bank's demands:

$$X_{B,t}^* \leq \overline{X}_{B,t} = \begin{cases} 0 & \text{if } \Delta X_{N,t} \leq 4M_L \\ \frac{1}{4}\Delta X_{N,t} - 2M_L + 4M_L^2\left(\frac{1}{\Delta X_{N,t}}\right) & \text{if } \Delta X_{N,t} \geq 4M_L \end{cases}. \tag{18}$$

This bound is intuitive. If the arbitrageur is unconstrained (i.e., $\Delta X_{N,t} \leq 4M_L$), the bank cannot place any demands without making him worse off from communicating with her. When the arbitrageur is constrained, the bank has some freedom to front-run, but only to a limited extent. Likewise, the bank's IC constraint, along with:

$$\Pi_{B,t}^c = \frac{\lambda}{2}\Delta X_{N,t} \cdot X_{B,t}^*, \tag{19}$$

implies a lower bound on the bank's demands:

$$X^*_{B,t} \geq \underline{X}_{B,t} = \begin{cases} \max\left\{0, \frac{1}{8}\Delta X_{N,t} - \frac{2V_t^*}{\lambda}\left(\frac{1}{\Delta X_{N,t}}\right)\right\} & \text{if } \Delta X_{N,t} \leq (4+2\sqrt{2})M_L \\ \\ \max\left\{0, \frac{1}{4}\Delta X_{N,t} - M_L + \right. & \text{if } \Delta X_{N,t} \geq (4+2\sqrt{2})M_L \\ \left.\left(M_L 2 - \frac{2V_t^*}{\lambda}\right)\left(\frac{1}{\Delta X_{N,t}}\right)\right\} & \end{cases} . \tag{20}$$

This lower bound indicates that if the bank's temptation to deviate is high enough, she must be allowed to trade a certain amount or else she'll choose to deviate.

These two bounds have powerful implications on the possibility of obtaining capital from the bank. Communication between the arbitrageur and the bank is rational if and only if both the arbitrageur's IR and the bank's IC constraints are satisfied (i.e., $\overline{X}_{B,t} \geq \underline{X}_{B,t}$). Proposition 6 shows that this is not always possible.

Proposition 6: *Information revelation by the arbitrageur cannot be sustained if* $\Delta X_{N,t} > \delta x_t^*$ *where:*

$$\delta x_t^* = \begin{cases} 0 & \text{if } V_t^* \leq \lambda M_L^2 \\ \\ 8M_L - 4\sqrt{2M_L^2 - \frac{V_t^*}{\lambda}} & \text{if } \lambda M_L^2 \leq V_t^* \leq \left(\frac{1+2\sqrt{2}}{2}\right)\lambda M_L^2 \\ \\ 3M_L + \frac{2V_t^*}{\lambda}\left(\frac{1}{M_L}\right) & \text{if } V_t^* \geq \left(\frac{1+2\sqrt{2}}{2}\right)\lambda M_L^2 \end{cases} . \tag{21}$$

This result is illustrated in Figure 2:

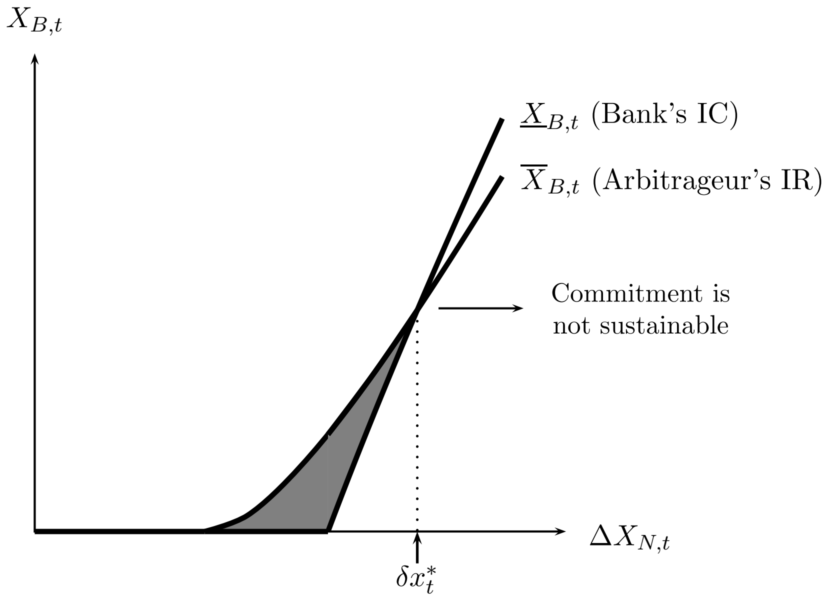

Figure 2: Illustration of Proposition 6.

A sketch of the reasoning behind this result can be stated in a straightforward manner. For large enough mispricing (i.e., $\Delta X_{N,t} \geq (4+2\sqrt{2})M_L$), the arbitrageur is constrained even when the bank deviates. Since the bank is unconstrained, she induces a price spread that is half the one that would obtain if the arbitrageur didn't reveal his information. Therefore, in the event of deviation, the arbitrageur loses exactly half of the profits he would have earned if he chose $R = 0$. This loss is roughly the bank's gain, and since the arbitrageur's outside option is unbounded, its value will eventually become greater than the bank's reputation value V_t^*. In other words, stealing about half the arbitrageur's profits can be better than holding on to future business so the bank cannot commit to improving upon the arbitrageur's outside option. In this case, communication between the two breaks down and arbitrage is limited like in the one-shot game.

This is the main result of the paper. Financial relationships are limited in their ability to mitigate the hold-up problem between arbitrageurs and banks. More importantly, they are insufficient when mispricing is large so that reputation-based commitment fails when it is needed most for arbitrage. It is important to note that this result cannot be obtained in a standard repeated game with $\Delta X_{N,t}$ held constant across periods. In fact, performing comparative statics on this alternative framework is likely to produce the opposite conclusion that arbitrage is limited when $\Delta X_{N,t}$ is small. To see this, look at the bank's incentive compatibility constraint under constant mispricing:

$$\frac{\lambda}{8}\Delta X_N^2 - \frac{\lambda}{2}M_L\Delta X_N + \frac{\lambda}{2}M_L^2 \leq \frac{1}{1-\delta}\left[\frac{\lambda}{8}\Delta X_N^2 - \lambda M_L \Delta X_N + 2\lambda M_L^2\right].^7 \tag{22}$$

This reduces to:

$$\left[\frac{\lambda\delta}{8}\right]\Delta X_N^2 - \left[\frac{\lambda(1+\delta)}{2}\right]M_L\Delta X_N + \left[\frac{\lambda(3+\delta)}{2}\right]M_L^2 \geq 0. \tag{23}$$

Though this IC constraint may be violated, it is guaranteed to hold for sufficiently high values of ΔX_N.

The main difference between this setting and ours is that allowing ΔX_N to be time-varying creates a wedge between the bank's deviation profit and the value of financial relationships by distinguishing between factors that affect *current* and *future* mispricing. This ensures that when current mispricing increases, there is not a corresponding (and amplified) increase in relationship values. As a result, bank commitment becomes more difficult to sustain in times where $\Delta X_{N,t}$ is high.

Some front-running is observed in our equilibrium but its degree is always acceptable to the arbitrageur. Myopically, the bank could do more to hold-up the arbitrageur, but she finds it in her interests to restrain herself. Furthermore, this degree varies with the attractiveness of the arbitrage. The arbitrageur needs to allow the bank to behave more opportunistically when these opportunities are greatest because it is in those events that the temptation to deviate is strongest. In this sense, we isolate a phenomenon that is broadly consistent with what Abolafia (Ch. 1, 1996) refers to as "cycles of opportunism" in financial markets.

A second element of optimal collusion that completes our characterization of the efficient equilibria:

Lemma 7: *An efficient SPE satisfies:*

$$R_t^* = \begin{cases} 0 & \text{if } \Delta X_{N,t} \leq 4M_L \\ 1 & \text{if } 4M_L \leq \Delta X_{N,t} \leq \delta x_t^* \\ 0 & \text{if } \Delta X_{N,t} \geq \delta x_t^* \end{cases}. \tag{24}$$

[7] To simplify exposition, we assume that $\Delta X_N \geq 4(2+\sqrt{2})M_L$ and that the bank keeps all the surplus from its relationship with the arbitrageur.

This result states that the bank's commitment ability is monotonic. If it can make an adequate commitment at a given level of mispricing, it can also do so at lower levels of mispricing. Lemma 7 also implies that efficient equilibria require the arbitrageur and the bank to collude whenever possible.

We can summarize the characterization of the equilibria to obtain:

Corollary 8: *A strategy profile is an efficient equilibrium if and only if:*

$$(R_t^*, X_{A,t}^*) = \begin{cases} (0, \frac{1}{4}\Delta X_{N,t}) & \text{if } \Delta X_{N,t} \leq 4M_L \\ \\ (1, M_t^*) & \text{if } 4M_L \leq \Delta X_{N,t} \leq \delta x_t^* \\ \\ (0, M_L) & \text{if } \Delta X_{N,t} \geq \delta x_t^* \end{cases} . \tag{25}$$

and

$$(M_t^*, X_{B,t}^*) = \begin{cases} (M_L, 0) & \text{if } \Delta X_{N,t} \leq 4M_L \\ \\ (\frac{1}{4}\Delta X_{N,t} - X_{B,t}, X_{B,t}) & \text{if } 4M_L \leq \Delta X_{N,t} \leq \delta x_t^* \\ \\ (M_L, 0) & \text{if } \Delta X_{N,t} \geq \delta x_t^* \end{cases} . \tag{26}$$

where $X_{B,t} \in [\underline{X}_{B,t}, \overline{X}_{B,t}]$.

Furthermore, equilibrium price spreads follow immediately from Corollary 8:

Corollary 9: *Equilibrium price spreads are given by:*

$$\Delta p_t^* = \begin{cases} \frac{\lambda}{2}\Delta X_{N,t} & \text{if } \Delta X_{N,t} \leq \delta x_t^* \\ \\ \lambda\Delta X_{N,t} - 2\lambda M_L & \text{if } \Delta X_{N,t} \geq \delta x_t^* \end{cases} . \tag{27}$$

This is illustrated in Figure 3:

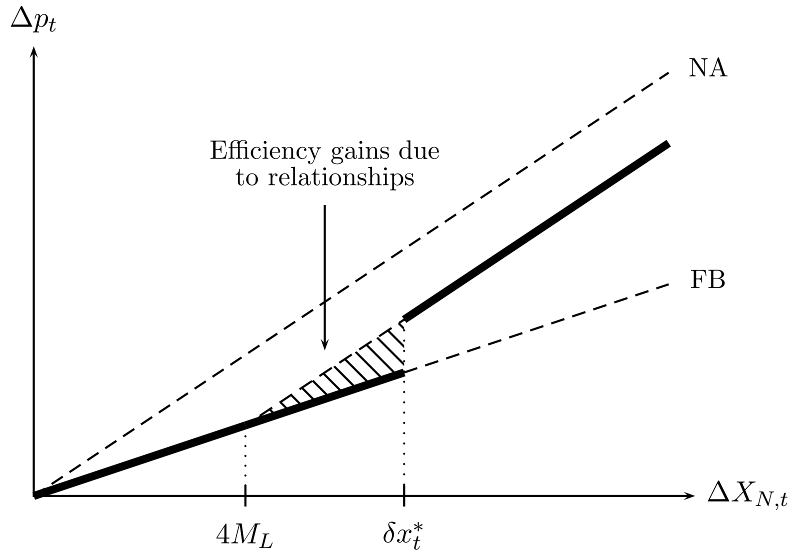

Figure 3: Equilibrium price spreads.

As remarked earlier, whenever there is communication between the arbitrageur and the bank, optimal collusion implies that price spreads are as in the first-best equilibrium. When the channel for informable finance breaks down though, they equal those of the one-shot game and arbitrage only has a fixed effect in correcting mispricing. Figure 3 illustrates the benefits of financial relationships on arbitrage. The thick line denotes the price spread in efficient SPEs of the repeated game. The shaded region represents the gains from relationships. This region becomes larger the more valuable a bank's reputation. An interesting feature of these equilibria is that small differences in initial mispricing, say $\Delta X_{N,t} = \delta x_t^*$ versus $\Delta X_{N,t} = \delta x_t^* + \epsilon$, can lead to discontinuous changes in equilibrium mispricing. This is because relationships have their greatest effect in reducing mispricing at $\Delta X_{N,t} = \delta x_t^*$ and no effect on price spreads once communication breaks down.

Under fairly general conditions, the limited arbitrage problem becomes worse as the arbitrageur's initial wealth increases. To simplify the exposition of this result, we make the weak assumption that the surplus allocation rule between the arbitrageur and the bank is independent of M_L.[8] In line with the literature on Nash bargaining, surplus is given by the difference between first-best profits and the sum of the arbitrageur's outside option and the bank's net payoff from deviation:

$$S_t \equiv \Pi_{M,t} - \left[\overline{\Pi}_{A,t} + \left(\Pi^d_{B,t} - V_t\right)\right]. \tag{28}$$

[8]If anything, we expect that the arbitrageur's bargaining power relative to the bank would increase with M_L. It can be shown that our results in Proposition 10 continue to hold in this case.

The above assumption implies that there is a process α_t that does not depend on M_L and satisfies:

$$\Pi^c_{B,t} = \left(\Pi^d_{B,t} - V_t\right) + \alpha_t \cdot S_t \tag{29}$$

$$= \alpha_t \cdot \left(\Pi_{M,t} - \overline{\Pi}_{A,t}\right) + (1 - \alpha_t) \cdot \left(\Pi^d_{B,t} - V_t\right). \tag{30}$$

This result is illustrated in Proposition 10:

Proposition 10: *Let $\delta^*_t(X)$ denote the value of δ^*_t in an equilibrium with $M_L = X$. Assuming that $0 < M'_L < M''_L$ and $\delta^*_t(M'_L) > 0$, it follows that $\delta^*_t(M'_L) > \delta^*_t(M''_L)$.*

Higher arbitrageur wealth has two effects. First, the bank gets lower profits following communication because the arbitrageur's outside option increases. Second, the likelihood that the arbitrageur is constrained and in need of the bank's services decreases. As a result, the value of financial relationships falls when M_L increases and arbitrage becomes more limited.

4 Refinement and Examples

Our characterization of efficient equilibria allows for many different possible outcomes because we have not made enough assumptions about how surplus is split between the arbitrageur and the bank.[9] In this section, we present a simple refinement of our equilibrium, based on competition between banks in forming relationships with arbitrageurs, that pins down this allocation of surplus and narrows our set of equilibria down to a unique one.

4.1 Simple Refinement

Consider an extension of our stage game that allows banks to make offers of relational contracts to the arbitrageur after he receives his private signal. Such promises prescribe a contingent action plan for the bank as a function of $\Delta X_{N,t}$. This implicitly defines a surplus allocation rule between the bank and the arbitrageur. We consider two extreme cases of competition between banks. In the first case, which we refer to as monopoly, the bank is alone in making an offer.[10] In the second case, there are a large number of banks who induce perfect competition in bidding.

[9] With the exception of Proposition 10, we've allowed $X^c_{B,t}$ to be any curve in the shaded area of Figure 2 (i.e. anything between $\underline{X}_{B,t}$ and $\overline{X}_{B,t}$).

[10] Using the term monopoly here is a bit abusive since we've motivated some earlier results with the assumption that the arbitrageur can start up new relationships with other equally proficient banks.

In the monopoly case, the bank gets all the surplus from the relationship. As a result, if the arbitrageur and the bank cooperate, the latter receives the first-best level of profits minus the arbitrageur's outside option. This implies a unique efficient SPE that is characterized by V_t^*. This relationship value is given by the largest solution to the fixed-point problem[11]:

$$V_t^* = \frac{\delta}{1-\delta} \int_0^{\delta x_{t,M}^*} \Pi_{B,t}^{c,M}(\Delta X_{N,t}) dF(\Delta X_{N,t}) \tag{31}$$

where $\Pi_{B,t}^{c,M}$ is defined in Lemma 11 in the Appendix. Existence of the unique solution, V_M^*, is guaranteed since zero is always a solution to the equation. Further, V_M^* is independent of time so the resulting equilibrium is stationary.

Under perfect competition, the bank promises the arbitrageur as much surplus as she possibly can. In other words, she either sets $X_{B,t}^* = 0$ or, when she does trade, binds herself to her IC constraint by setting:

$$\Pi_{B,t}^{c,PC} + V_t^* = \Pi_{B,t}^d.$$

As in the monopoly case, there is a unique efficient SPE characterized by V_t^* which is the largest solution to the problem:

$$V_t^* = \frac{\delta}{1-\delta} \int_0^{\delta x_{t,PC}^*} \Pi_{B,t}^{c,PC}(\Delta X_{N,t}) dF(\Delta X_{N,t}). \tag{32}$$

where $\Pi_{B,t}^{c,PC}$ is defined in Lemma 12 in the Appendix. This produces a solution, V_{PC}^*, that is also independent of time.

In both environments, stationarity is driven by the independence assumption on $\Delta X_{N,t}$ and the fixed surplus allocation rule. It also requires that arbitrageurs be short-lived. This is a different foundation than found elsewhere in the relational contracting literature. For example, the weak stationarity lemma in Levin (2003) is driven by the unlimited wealth of the principal and agent which allows for settling up on a period-by-period basis. Nevertheless, it should be noted that stationarity of V is not important for the main results of this paper.

[11] It is easy to verify that, if there are two solutions to the fixed-point problem with values V_t^* and $V_t^{**} > V_t^*$, moving from the equilibrium implied by V_t^* to V_t^{**} is an interim Pareto improvement. Therefore, only the equilibrium with V_t^{**} is an efficient SPE.

Communication is easier to sustain in the monopoly case than the perfect competition case, since aggressive competition between banks forces them to bid away their share of the surplus and lowers their valuation of financial relationships. This worsens their ability to make commitments. Proposition 13 formalizes this intuition:

Proposition 13: *Comparing the efficient SPE from the monopoly and perfect competition cases, we have that:*

$$V_M^* \geq V_{PC}^* \tag{33}$$

and

$$\delta x_M^* \geq \delta x_{PC}^*. \tag{34}$$

4.2 Binomial Distribution Example

Assume that $\Delta X_{N,t}$ equals $\epsilon > 0$ with probability p and 0 otherwise. Let $M_L = \beta\epsilon/4$ with $\beta < 1$ so that the arbitrageur is constrained when there is mispricing. The smaller the value of β, the more severely constrained is the arbitrageur when $\Delta X_{N,t} = \epsilon$.

When the bank has monopoly power, we can solve for the efficient equilibrium by assuming that information revelation can be sustained at $\Delta X_{N,t} = \epsilon$ and checking for consistency afterwards. If there is communication, we have:

$$V_M^* = \left(\frac{\delta}{1-\delta}\right) E\left[\Pi_{FB} - \overline{\Pi}_A\right] = \frac{\lambda p}{8}\left(\frac{\delta}{1-\delta}\right)(1-\beta)^2\epsilon^2. \tag{35}$$

We check for consistency by verifying whether the bank's IC constraint is satisfied. It is if:

$$\beta \leq \overline{\beta}_M = \frac{4\theta + 2 - 2\sqrt{\theta+1}}{4\theta+3} \tag{36}$$

where $\theta = p\delta/(1-\delta)$. If the IC constraint doesn't hold, V_M^* equals zero, since no relational contract is enforceable.

Notice that $\overline{\beta}_M$ is monotonically increasing in θ. This implies that there are no limits to arbitrage if mispricing is sufficiently frequent or the bank is patient enough. This is intuitive because both lead to higher relationship values for the bank. What may be surprising is that λ and ϵ do not affect $\overline{\beta}_M$. This is due to scaling effects: all profit functions are perfectly linear in λ and quadratic in ϵ, so the arbitrageur's IR and the bank's IC constraints are unaffected by these parameter value. In the case of ϵ, the scaling argument requires us to hold β fixed because of the arbitrageur's outside option.

Likewise, in the case of perfect competition, if communication between the arbitrageur and the bank occurs at $\Delta X_{N,t} = \epsilon$, we have:

$$V_{PC}^{*} = \frac{\lambda\theta\left(\frac{1}{8} - \frac{\beta}{8} + \frac{\beta^2}{32}\right)\epsilon^2}{1+\theta}. \tag{37}$$

This can be sustained if the arbitrageur's IR constraint holds. That is, if:

$$\beta \leq \overline{\beta}_{PC} = \frac{4\theta + 2 - 2\sqrt{\theta+1}}{4\theta+3} \tag{38}$$

The same results and intuitions from the monopoly case hold here. Surprisingly, we have that $\overline{\beta}_{PC} = \overline{\beta}_{M}$ in this setting. This is a particular feature of the binomial distribution example. When there is only one level of possible mispricing, competition does not matter at the margin where arbitrage becomes limited because there is no surplus, beyond the bank's required rent from her IC constraint, to bargain over. To investigate the effect of competition between banks on the limits to arbitrage problem requires multiple potential levels of mispricing in order to produce regions where the division of surplus matters. This motivates our next example.

4.3 Normal Distribution Example

We assume that $X_{N1,t}$ and $X_{N2,t}$ are joint normally distributed so that we can write $f(\Delta X_{N,t}) = 2 \cdot n(\Delta X_{N,t}; \sigma^2)$ where $n(.;\sigma^2)$ is the density of a normal distribution with mean 0 and standard σ. We also set $\beta = 4M_L/\sigma$. This parameter can be interpreted as the standard deviation shock to $\Delta X_{N,t}$ that is required to make the arbitrageur constrained under his default financial constraint. This is economically similar to the β parameter from the binomial distribution case. We also write $\langle \delta x^* \rangle \equiv \delta x^*/\sigma$. This is the threshold where arbitrage becomes limited normalized by the standard deviation of mispricing. Holding $\langle \delta x^* \rangle$ constant fixes the frequency of limited arbitrage in this setting.

As in the binomial case, the threshold point is invariant to scaling, as shown in the following Lemma:

Lemma 14: *In both the monopoly and perfect competition cases, holding β constant, $\langle \delta x^* \rangle$ does not depend on σ and λ.*

This example cannot be solved in closed form, so we provide numerical solutions. It is useful to define $c = 2 \cdot N(-\beta)$ where $N(.)$ is the cumulative distribution for a standard normal. This value represents the percentage of arbitrage opportunities where the arbitrageur is constrained under his default financial constraint. Figure 4 illustrates the value of $\langle \delta x^* \rangle$ as a function of c and the discount factor δ in both the monopoly and perfect competition cases:

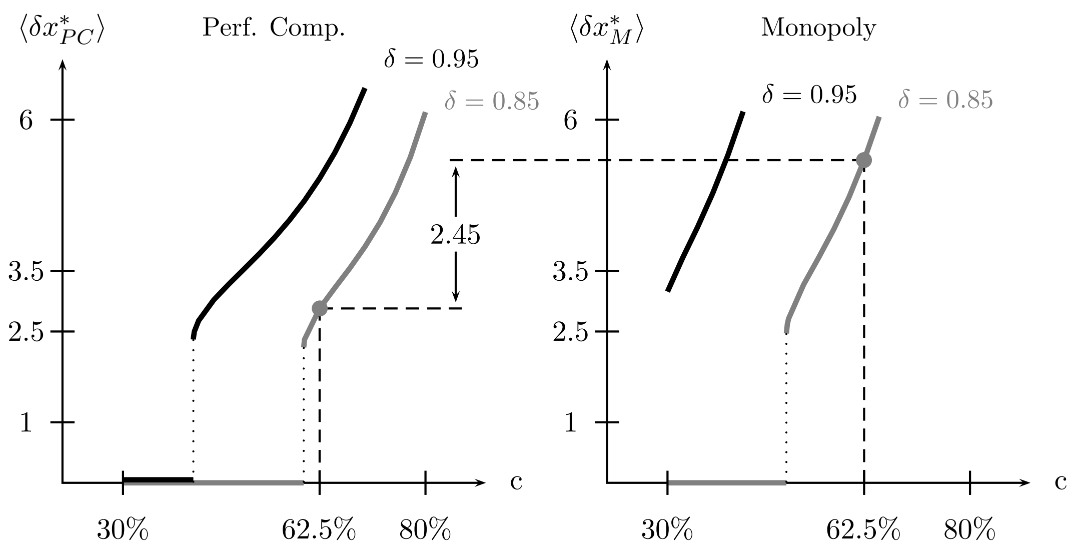

Figure 4: Relationship thresholds.

Notice that there is a substantial difference in the thresholds between the perfect competition and monopoly cases. Consistent with Proposition 13, we observe limited arbitrage more frequently in the case of perfect competition among banks. For instance, as illustrated in the figure, when $\delta = 0.85$ and the arbitrageur is constrained in 62.5% of the arbitrage opportunity cycles, limited arbitrage is over 40,000 times more likely in the case of perfect competition ($2 \cdot N(-2.88)$ vs. $2 \cdot N(-5.33)$). The frequency of limited arbitrage is also economically significant. If the arbitrageur is sufficiently wealthy (low c), then financial relationships have no value and outcomes are as in the one-shot game. This is expected since the arbitrageur rarely needs the bank and has a high outside option, thus producing a low value to reputation that the bank is unable to ever meet the arbitrageur's outside option.

Perhaps surprisingly, Figure 4 also indicates that a small change in c can have a large and discontinuous impact on the role that relationships play in funding arbitrage activity. This is due to the arbitrageur's outside option. As he becomes wealthier, he requires a larger share of profits, which lowers the value of relationships to the bank. This reduction in value also lowers δx^*, which further lowers V^*. The discontinuity arises, because $V^* \geq \lambda M_L^2$ is needed to satisfy the bank's IC constraint at $\Delta X_{N,t} = 4M_L$.

5 Additional Extensions and Discussion

5.1 Explicit Contracting

In this section, we consider a special case of explicit contracting, namely the possibility of signing labor contracts with banks. A labor contract specifies a non-negative and non-decreasing wage payment that is conditional on profits and possibly a trading budget that sets an upper bound on the position, $\overline{M}$, the bank employee can undertake. However, the content of communication between the employee and the bank is still non-contractible. Employment contracts can be renegotiated once an employee joins the bank and the bank maintains her ability to attract employees even if she deviates from her implicit agreements with arbitrageurs who trade on their own.

There are two types of risk-neutral agents: arbitrageurs, identical to the ones from previous sections, and speculators. A proportion θ of potential employees are speculators and the cumulative distribution function of $\Delta X_{N,t}$ is monotonically increasing. Speculators have no private information and can only make risky investments that generate mean zero profits with distribution G per unit. Due to their limited liability constraint, they may extract rents from the bank by risk-shifting. Speculators also have access to an alternative employment opportunity that pays a fixed wage $\overline{w} > 0$. Arbitrageurs' outside opportunity is to trade on their own. For simplicity, we will assume that the bank is a monopolist and θ is sufficiently close to 1.[12] The latter induces a fly-by-night constraint where the bank chooses to screen out all speculators. The optimal contract between the bank and its employees is given in Proposition 15:

Proposition 15: *Let Ω denote the set of arbitrageur types $\Delta X_{N,t}$ that are hired by the bank. We have that: (i) the optimal employment contract specifies a finite upper bound $\overline{M}$, (ii) $\Omega \subseteq [M_L/4, \overline{M}/4]$, (iii) the speculators' screening condition binds:*

$$\overline{w} = \max_{X \leq \overline{M}} E^G[W(X \cdot \tilde{\Pi})], \tag{39}$$

(iv) the wage offer is given by:

[12] This assumption is made for expositional reasons and the results continue to hold if $\theta \in (0, 1)$.

$$W(\Pi) = \begin{cases} 0 & \text{if } \Pi \leq 2\lambda M_L^2 \\ \\ 2M_L\sqrt{2\lambda\Pi} - 2\lambda M_L^2 & \text{if } 2\lambda M_L^2 \leq \Pi \leq 2\lambda \overline{M}^2 \text{ and } 2\sqrt{\frac{2\Pi}{\lambda}} \in \Omega \\ \\ \sup_{\hat{\Pi}<\Pi} W(\hat{\Pi}) & \text{if } 2\lambda M_L^2 \leq \Pi \leq 2\lambda \overline{M}^2 \text{ and } 2\sqrt{\frac{2\Pi}{\lambda}} \notin \Omega \\ \\ 4\lambda M_L \overline{M} - 2\lambda M_L^2 & \text{if } \Pi \geq 2\lambda \overline{M}^2 \end{cases}, \tag{40}$$

and (v) Arbitrage is still limited.

The main parts of this proposition are (i) and (v). Both follow from the fact that the fly-by-night constraint induces the bank to limit the rents speculators can extract. In order to do this, the bank cannot leave both the maximal wage and trading budget in its employment offer unbounded. The upper bound on wages, which is a function of $\overline{M}$, follows from the requirement that the contract be renegotiation proof. Unfortunately, these bounds also screen out the arbitrageurs with the most profitable arbitrage opportunities. Hence, arbitrage activity is still limited. It should be noted that the form of contract obtained here is similar to those observed in practice. For instance, proprietary traders employed by investment banks are often subject to position limits and receive bonuses that depend on their trading profits.

Interestingly, allowing for explicit contracts can increase the frequency of limited arbitrage. This is due to the fact that, under explicit contracting, the bank has less to lose from deviation because it keeps its profits from continuing to employ proprietary traders.

Proposition 16: *If $\delta x_M^* > 4\overline{M} > 4M_L$, then allowing explicit contracts leads to more severe limits to arbitrage.*

The observation that explicit contracts can crowd out implicit ones is also emphasized, in a managerial compensation setting, by Baker-Gibbons-Murphy (1994).

5.2 The Market Structure of Arbitrage

In addition to investment banks, several other organizations, like fund-of-funds and seeders, have emerged in the market for funding arbitrageurs. Fund-of-funds invest in a portfolio of established hedge funds while seeders help emerging managers obtain initial capital to get up and running. These investors have significantly more expertise than most investors and usually negotiate favorable liquidity treatment from the funds they invest in. Therefore, they can play the role of informable financier for broad strategy at a fund's initiation as well as temporary investment opportunities that arise during the course of operations.

Furthermore, without their own trading desks, these organizations cannot directly implement the arbitrageur's strategy. They can only indirectly expropriate by sharing information about the arbitrageur's profit opportunities with other funds they invest in. This is similar to the hold-up problem investigated by Cestone-White (2003). Interestingly, consistent with the optimal contracting solution derived to overcome this problem, fund-of-funds and seeders often acquire equity-like claims in the hedge fund's *management company.*

The existence of these institutions is consistent with certain implications of our theory. Namely, that there are benefits to setting up institutions that can serve as informable financiers without having a substantial ability to expropriate the arbitrageur. However, it must be noted that investment banks still provide the bulk of financing to arbitrageurs. There are several potential explanations for this fact. The industry may still be maturing and facing its own wealth constraints. Alternatively, renegotiation of margin agreements and the reallocation of margin use across strategies may be less costly and time consuming than obtaining additional finance from fund-of-funds and seeders. This would explain why investment banks are especially dominant as a source of short-term financing. The industry may also be less effective in understanding and monitoring the arbitrageurs' strategies because of its lack of direct participation in trading.

Regarding the organizational structure of banks, there is a debate on whether Chinese walls should be required between prime brokerage and proprietary trading desks. Our model illustrates one benefit of the *absence* of Chinese walls, namely that communication between divisions leads to improved arbitrage when prime brokers lack the expertise to understand the strategies themselves. However, these walls will be useful whenever the prime broker can understand the arbitrageur because the expropriation threat is absent when he is prohibited from leaking information to proprietary traders. Therefore, the optimality of enforcing Chinese walls between the two divisions depends on a trade-off between these factors.

5.3 Financing Innovation

Our model can also be applied to the more general setting of financing innovation. Consider a setting where a sequence of entrepreneurs want to fund projects with public debt and private equity. This is isomorphic to our model in sections 2 through 4, where $\Delta X_{N,t}$ now represents the profitability of a project.[13] Let M_L denote the production capacity that the entrepreneur can achieve while relying only on arms-length debt financing. The entrepreneur can also obtain capital through a relationship with an informable financier, as in our previous model. Likewise, the extension of section 5.1 can be viewed as allowing for other types of securities, like equity, to be issued when raising capital without disclosing private information.

An interpretation of the informable financier in this context would be a venture capitalist, who has expertise in specific industries and can gauge the quality of a project, but also has the contacts to implement the project herself. As suggested in the quote below, reputation is an important concern here:

> *We are extremely conscious that corporations and entrepreneurial ventures can be in conflict... if [the entrepreneurs] don't know us or have never interacted with Cisco, there is an initial concern that needs to be overcome... We've overcome such concerns by building a track record. We have enough references within the venture capital community that we can say, 'Hey, why don't you talk to John Doerr or go to Don Valentine and ask them what they think about having Cisco as an investor.' Pretty unanimously we get over the hurdle.*
>
> - Mike Volpi, Head of Cisco Ventures (in Gupta p.32)

However, our model suggests that reputation alone cannot always guarantee first-best implementation. The most profitable projects will still be undercapitalized.

6 Conclusion

We provide an economic foundation for the limits to arbitrage problem, based on a hold-up problem between arbitrageurs and informable financiers. Reputational concerns on the part of the financiers partially mitigate this problem. However, relationships fail when needed most. Additionally, more bargaining power in the hands of arbitrageurs and increased competition among banks can make everyone worse off since it reduces the value of relationships to the bank. Holding the degree of bargaining power fixed, we also show that higher initial arbitrageur wealth worsens the effectiveness

[13]This can explicitly be viewed as a consumer demand shock in the spirit of Rotemberg-Saloner (1986).

of informable finance. Finally, allowing for explicit contracts does not necessarily alleviate the limits to arbitrage problem and, in fact, can worsen it.

7 Appendix

7.1 Proofs

This section contains the unproven propositions from sections 3 to 5 of the paper.

Proof of Proposition 2: We proceed by backward induction. If $R = 0$, it follows by assumption that $M^* = M_L$ and $X_B^* = 0$ and from Proposition 1 that $X_A^* = \min(X_A^{FB}, M_L)$. Therefore,

$$X_A^* = \begin{cases} \frac{1}{4}\Delta X_N & \text{if } R = 0 \text{ and } \Delta X_N \leq 4M_L \\ \\ M_L & \text{if } R = 0 \text{ and } \Delta X_N \geq 4M_L \end{cases}, \tag{41}$$

and

$$\Delta p^* = \begin{cases} \frac{\lambda}{2}\Delta X_N & \text{if } R = 0 \text{ and } \Delta X_N \leq 4M_L \\ \\ \lambda \Delta X_N - 2\lambda M_L & \text{if } R = 0 \text{ and } \Delta X_N \geq 4M_L \end{cases}. \tag{42}$$

Payoffs are given by:

$$\Pi_A^* = \begin{cases} \frac{\lambda}{8}\Delta X_N^2 & \text{if } R = 0 \text{ and } \Delta X_N \leq 4M_L \\ \\ \lambda M_L \Delta X_N - 2\lambda M_L^2 & \text{if } R = 0 \text{ and } \Delta X_N \geq 4M_L \end{cases}. \tag{43}$$

Now assume that $R = 1$. Given ΔX_N and X_B, the arbitrageur chooses his demand such that:

$$\max_{X_A \leq M_L} \lambda X_A \left(\Delta X_N - 2X_A - 2X_B\right). \tag{44}$$

This implies that:

$$X_A^* = \begin{cases} \frac{1}{4}\Delta X_N - \frac{1}{2}X_B & \text{if } R = 1 \text{ and } X_B \geq \frac{1}{2}\Delta X_N - 2M_L \\ \\ M_L & \text{if } R = 1 \text{ and } X_B \leq \frac{1}{2}\Delta X_N - 2M_L \end{cases}. \tag{45}$$

As a result, given ΔX_N, the bank's payoff as a function of its own choice of trade is:

$$\Pi_B(X_B) = \begin{cases} -2\lambda \left(X_B - \frac{1}{4}\Delta X_N + \frac{1}{2}M_L\right)^2 + \\ \frac{\lambda}{8}\Delta X_N^2 - \frac{\lambda}{2}M_L\Delta X_N + \frac{\lambda}{2}M_L^2 & \text{if } R = 1 \text{ and } X_B \leq \frac{1}{2}\Delta X_N - 2M_L \\ \\ -\lambda \left(X_B - \frac{1}{4}\Delta X_N\right)^2 + \frac{\lambda}{16}\Delta X_N^2 & \text{if } R = 1 \text{ and } X_B \geq \frac{1}{2}\Delta X_N - 2M_L \end{cases}. \tag{46}$$

This piecewise quadratic objective consists of two parts. For small values of X_B, the arbitrageur is constrained while, for higher values of X_B, he is unconstrained. The quadratic function corresponding to the constrained segment reaches its maximum at $X_B = \frac{1}{4}\Delta X_N - \frac{1}{2}M_L$ while the unconstrained version achieves its maximum at a strictly higher point, namely $X_B = \frac{1}{4}\Delta X_N$. We now determine the bank's optimal X_B^* given $R = 1$ and a particular ΔX_N. To do so, we consider 3 regions of ΔX_N:

Region 1: $\Delta X_N \leq 6M_L$. We have $\frac{1}{4}\Delta X_N - \frac{1}{2}M_L \geq \frac{1}{2}\Delta X_N - 2M_L$ which implies that there is only one local/global maximum in the bank's objective function, namely $X_B^* = \frac{1}{4}\Delta X_N$. Notice that this implies the arbitrageur is unconstrained (i.e. $X_A \leq M_L$).

Region 2: $\Delta X_N \geq 8M_L$. We have $\frac{1}{4}\Delta X_N \leq \frac{1}{2}\Delta X_N - 2M_L$ and, therefore, the only local/global maximum in the bank's objective is $X_B^* = \frac{1}{4}\Delta X_N - \frac{1}{2}M_L$. The arbitrageur is constrained.

Region 3: $6M_L \leq \Delta X_N \leq 8M_L$. In this region, there are two local maxima in the bank's objective: $X_B = \frac{1}{4}\Delta X_N$ (constrained) and $X_B = \frac{1}{4}\Delta X_N - \frac{1}{2}M_L$ (unconstrained). The global maximum is given by the constrained case if and only if

$$\frac{\lambda}{8}\Delta X_N^2 - \frac{\lambda}{2}M_L\Delta X_N + \frac{\lambda}{2}M_L^2 \geq \frac{\lambda}{16}\Delta X_N^2, \tag{47}$$

or equivalently,

$$\Delta X_N \geq (4 + 2\sqrt{2})M_L. \tag{48}$$

Therefore, we have:

$$X_B^* = \begin{cases} \frac{1}{4}\Delta X_N & \text{if } R = 1 \text{ and } \Delta X_N \leq (4+2\sqrt{2})M_L \\ \\ \frac{1}{4}\Delta X_N - \frac{1}{2}M_L & \text{if } R = 1 \text{ and } \Delta X_N \geq (4+2\sqrt{2})M_L \end{cases} \tag{49}$$

and

$$X_A^* = \begin{cases} \frac{1}{8}\Delta X_N & \text{if } R = 1 \text{ and } \Delta X_N \leq (4+2\sqrt{2})M_L \\ \\ M_L & \text{if } R = 1 \text{ and } \Delta X_N \geq (4+2\sqrt{2})M_L \end{cases}. \tag{50}$$

Meanwhile, price spreads and arbitrageur profits are given by:

$$\Delta p^* = \begin{cases} \frac{\lambda}{4}\Delta X_N & \text{if } R = 1 \text{ and } \Delta X_N \leq (4+2\sqrt{2})M_L \\ \\ \frac{\lambda}{2}\Delta X_N - \lambda M_L & \text{if } R = 1 \text{ and } \Delta X_N \geq (4+2\sqrt{2})M_L \end{cases} \tag{51}$$

and

$$\Pi_A^* = \begin{cases} \frac{\lambda}{32}\Delta X_N & \text{if } R = 1 \text{ and } \Delta X_N \leq (4+2\sqrt{2})M_L \\ \\ \frac{\lambda}{2}M_L\Delta X_N - \lambda M_L^2 & \text{if } R = 1 \text{ and } \Delta X_N \geq (4+2\sqrt{2})M_L \end{cases} \tag{52}$$

It is easy to verify that $\Pi_A^*(R = 0) \geq \Pi_A^*(R = 1)$ and, hence, $R^* = 0$. Q.E.D.

Corollary 3: *If the arbitrageur doesn't reveal his information to the bank, price spreads are given by:*

$$\Delta p^* = \begin{cases} \frac{\lambda}{2}\Delta X_N & \text{if } \Delta X_N \leq 4M_L \\ \\ \lambda\Delta X_N - 2\lambda M_L & \text{if } \Delta X_N \geq 4M_L \end{cases} . \tag{53}$$

Further, the arbitrageur and bank payoffs are:

$$\Pi_A^* = \begin{cases} \frac{\lambda}{8}\Delta X_N^2 & \text{if } R = 0 \text{ and } \Delta X_N \leq 4M_L \\ \\ \lambda M_L \Delta X_N - 2\lambda M_L^2 & \text{if } R = 0 \text{ and } \Delta X_N \geq 4M_L \end{cases} \quad \textit{and} \quad \Pi_B^* = 0. \tag{54}$$

Remark for Corollary 3. So long as the arbitrageur is unconstrained, the outcome is as in the first-best equilibrium. Otherwise, he binds to his default financial constraint and has a fixed convergence effect on the price spread (reducing it by $2\lambda M_L$).

Corollary 4: *If the arbitrageur does reveal his information to the bank, price spreads are given by:*

$$\Delta p^* = \begin{cases} \frac{\lambda}{4}\Delta X_N & \text{if } \Delta X_N \leq (4+2\sqrt{2})M_L \\ \\ \frac{\lambda}{2}\Delta X_N - \lambda M_L & \text{if } \Delta X_N \geq (4+2\sqrt{2})M_L \end{cases} . \tag{55}$$

Further, the arbitrageur and bank payoffs are:

$$\Pi_A^* = \begin{cases} \frac{\lambda}{32}\Delta X_N^2 & \text{if } R = 0 \text{ and } \Delta X_N \leq (4+2\sqrt{2})M_L \\ \\ \frac{\lambda}{2}M_L \Delta X_N - \lambda M_L^2 & \text{if } R = 0 \text{ and } \Delta X_N \geq (4+2\sqrt{2})M_L \end{cases} \tag{56}$$

and

$$\Pi_B^* = \begin{cases} \frac{\lambda}{16}\Delta X_N^2 & \text{if } R = 0 \text{ and } \Delta X_N \leq (4+2\sqrt{2})M_L \\ \\ \frac{\lambda}{8}\Delta X_N^2 - \frac{\lambda}{2}M_L\Delta X_N + \frac{\lambda}{2}M_L^2 & \text{if } R = 0 \text{ and } \Delta X_N \geq (4+2\sqrt{2})M_L \end{cases} , \tag{57}$$

respectively.

Remark for Corollary 4. When the initial mispricing is small, the arbitrageur is unconstrained following the bank's aggressive hold-up and the price spread is half that of the first-best equilibrium. Meanwhile, if $\Delta X_{N,t}$ is large enough, the arbitrageur is still constrained and the price spread equals half that of the one-shot game. In both cases though, equilibrium mispricing is smaller than in the first-best. This is due to the fact that the bank is unconstrained and doesn't internalize the arbitrageur's welfare (i.e. a standard competition effect).

Proof of Lemma 5: First of all, notice that if $X_{B,t} > 0$, then an unconstrained arbitrageur would choose $X_{A,t} > \frac{1}{4}\Delta X_{N,t} - X_{B,t}$ since his unwillingness to internalize the effect of his demands on the bank's profits would lead to total demands that exceed the first-best level. This implies that if the arbitrageur is constrained and $X_{B,t} + M_t = \frac{1}{4}\Delta X_{N,t}$, then $X_{A,t} = M_t$.

We now proceed to prove this lemma by contradiction. Assume there exists an efficient equilibrium such that $R_t^* = 1$ and $X_{B,t}^* + M_t^* \neq \frac{1}{4}\Delta X_{N,t}$ for some $\Delta X_{N,t}$. It follows that $\Pi_{A,t}^* + \Pi_{B,t}^* < \frac{\lambda}{8}\Delta X_{N,t}^2$. Consider an alternative set of strategies identical to this one with the exception that, at the aforementioned $\Delta X_{N,t}$, we have $X_{B,t}^{**} = \frac{2\Pi_{B,t}^*}{\lambda \Delta X_{N,t}}$ and $M_t^{**} = \frac{1}{4}\Delta X_{N,t} - X_{B,t}^{**}$. This implies that $\Pi_{A,t}^{**} + \Pi_{B,t}^{**} = \frac{\lambda}{8}\Delta X_{N,t} > \Pi_{A,t}^* + \Pi_{B,t}^*$ or equivalently, since $\Pi_{B,t}^{**} = \Pi_{B,t}^*$, that $\Pi_{A,t}^{**} > \Pi_{A,t}^*$. Notice that since the initial equilibrium satisfied both the arbitrageur's IR constraint and the bank's IC constraint, so does the new one (the IC constraint still holds since the value of the relationship V_t^{**} in the new construction is equal to the V_t^* from the initial equilibrium). This contradicts the claim that the SPE is efficient. Q.E.D.

Proof of Proposition 6: We consider three different regions for relationship values: (i) $V_t^* \leq \lambda M_L^2$, (ii) $\lambda M_L^2 \leq V_t^* \leq (1+2\sqrt{2})\lambda M_L^2/2$, and (iii) $V_t^* \geq (1+2\sqrt{2})\lambda M_L^2/2$.

First Region: If $V_t^* \leq \lambda M_L^2$, it is impossible to satisfy both the arbitrageur's IR and the bank's IC constraints at $\Delta X_{N,t} = 4M_L$. As a result, no communication can ever be sustained between the two and $\delta x_t^* = 0$.

Second Region: In this case, the breakdown of communication happens before $\Delta X_{N,t} = (4+2\sqrt{2})M_L$ and δx_t^* solves:

$$\frac{1}{4}\delta x_t^* - 2M_L + 4M_L^2\left(\frac{1}{\delta x_t^*}\right) = \frac{1}{8}\delta x_t^* - \frac{2V_t^*}{\lambda}\left(\frac{1}{\delta x_t^*}\right). \tag{58}$$

Third Region: When $V_t^* \geq (1+2\sqrt{2})\lambda M_L^2/2$, the breakdown occurs after $\Delta X_{N,t} = (4+2\sqrt{2})M_L$ and δx_t^* solves:

$$\frac{1}{4}\delta x_t^* - 2M_L + 4M_L^2\left(\frac{1}{\delta x_t^*}\right) = \frac{1}{4}\delta x_t^* - M_L + \left(M_L - \frac{2V_t^*}{\lambda}\right)\left(\frac{1}{\delta x_t^*}\right). \tag{59}$$

Q.E.D.

Proof of Lemma 7: We split this proof into two components: (i) show that $R_t^* = 1$ for $\Delta X_{N,t} = \delta x_t^*$, and (ii) show that if $R_t^* = 1$ for $\Delta X_{N,t}$ and $4M_L \leq \Delta X'_{N,t} \leq \Delta X_{N,t}$, then $R_t^* = 1$ for $\Delta X'_{N,t}$. Both are proven by contradiction.

First Component: Suppose there exists a Pareto-optimal equilibrium where $R_t^* = 0$ for $\Delta X_{N,t} = \delta x_t^*$. Consider a strategy profile that is identical to this equilibrium with the exception that it differs at the point $\Delta X_{N,t} = \delta x_t^*$. At that point, set $X_{B,t}' = \frac{2(\Pi_M,t-\overline{\Pi}_{A,t})}{\lambda \delta x_t^*}$ and $M_t' = \frac{2\overline{Pi}_{A,t}}{\lambda \delta x_t^*}$. It is easy to check that both the arbitrageur's IR and the bank's IC constraints are met for $\Delta X_{N,t} = \delta x_t^*$ and for all $\Delta X_{N,t}$ where $R_t^* = 1$. Therefore, the new strategy is a subgame perfect equilibrium that Pareto-dominates the aforementioned one. Contradiction.

Second Component: Assume that there exists a Pareto-optimal equilibrium where $R_t^* = 1$ for $\Delta X_{N,t}$ and $R_t^* = 0$ for $\Delta X_{N,t}'$. Since $R_t^* = 1$, we have $\overline{X}_t > \underline{X}_t$ for $\Delta X_{N,t}$. It follows that $\overline{X}_t' > \underline{X}_t'$ for $\Delta X_{N,t}'$ as well. Pick any $X_B \in [\underline{X}_t', \overline{X}_t']$ and consider a strategy profile that is identical to the stated equilibrium except that it sets $R_t' = 1$ and $X_{B,t} = X_B$ (while satisfying Lemma 5) for $\Delta X_{N,t}'$. It is easily verified that both the arbitrageur's IR and the bank's IC constraints are satisfied for $\Delta X_{N,t}'$ and for all $\Delta X_{N,t}$ where $R_t^* = 1$. This new strategy profile is subgame perfect and Pareto-dominates the aforementioned equilibrium. Contradiction. Q.E.D.

Proof of Proposition 10: We proceed to prove this by contradiction. Assume there exists two equilibria with $M_L' < M_L''$ such that $\delta_t^*(M_L') \leq \delta_t^*(M_L'')$. From Proposition 6, $\delta_t^*(M_L') < \delta_t^*(M_L'')$ implies that $V_t^*(M_L') < V_t^*(M_L'')$. Therefore, the bargaining power assumption implies that $\Pi_{B,t}^{c,*}(M_L') > \Pi_{B,t}^{c,*}(M_L'')$ in the interval $4M_L \leq \Delta X_{N,t} \leq \delta x_t^*(M_L')$. Now consider an alternative relational contract for M_L' with:

$$\Pi_{B,t}^{c,**}(M_L') = \begin{cases} \Pi_{B,t}^{c,*}(M_L') & \text{if } \Delta X_{N,t} \leq \delta_t^*(M_L') \\ \\ \Pi_{B,t}^{c,*}(M_L'') & \text{if } \Delta X_{N,t} > \delta_t^*(M_L') \end{cases} . \tag{60}$$

It follows that $V_t^{**}(M_L') > V_t^*(M_L'')$ and the bank's IC constraint is strictly satisfied up until $\delta_t^*(M_L'')$. Given optimal collusion, the arbitrageur's IR constraint is also met up till $\delta_t^*(M_L'')$ since his outside option is strictly lower when $M_L = M_L'$ than when $M_L = M_L''$. Continuity of $\Pi_{B,t}^d$ and the strict satisfaction of the bank's IC under $\Pi_{B,t}^{c,**}(M_L')$ implies that there exists an $\epsilon > 0$ such that we can construct a relational contract with:

$$\Pi_{B,t}^{c,***}(M_L') = \begin{cases} \Pi_{B,t}^{c,*}(M_L') & \text{if } \Delta X_{N,t} \leq \delta_t^*(M_L') \\ \\ \Pi_{B,t}^{c,*}(M_L'') & \text{if } \delta_t^*(M_L') < \Delta X_{N,t} \leq \delta_t^*(M_L'') \\ \\ \max_{\Delta X_{N,t}} \Pi_{B,t}^{c,*}(M_L'') & \text{if } \delta_t^*(M_L'') < \Delta X_{N,t} \leq \delta_t^*(M_L'') + \epsilon \\ \\ 0 & \text{if } \Delta X_{N,t} > \delta_t^*(M_L'') + \epsilon \end{cases} . \tag{61}$$

This contract satisfies $\delta x_t^{***}(M_L') > \delta x_t^*(M_L'')$ and contradicts the statement that the equilibrium with $\delta x_t^*(M_L')$ is efficient. Q.E.D.

Lemma 11: *A monopolist bank's cooperation profit, as a function of V_t^*, is given by:*

$$\Pi_{B,t}^{c,M} = \begin{cases} 0 & \text{if } \Delta X_{N,t} \leq 4M_L \\ \\ \frac{\lambda}{8}\Delta X_{N,t}^2 - \lambda M_L \Delta X_{N,t} + 2\lambda M_L^2 & \text{if } 4M_L \leq \Delta X_{N,t} \leq \delta x_t^* \\ \\ 0 & \text{if } \Delta X_{N,t} \geq \delta x_t^* \end{cases} \tag{62}$$

where δx_t^* depends on V_t^* as described in Proposition 6.

Proof of Lemma 11: When $V_t^* \leq \Pi_{B,t}^d$ and $\Delta X_{N,t} \leq \delta x_t^*$, the bank cannot commit to give the arbitrageur all the surplus, but can commit to cooperation. In this case, $\Pi_{B,t}^c = \Pi_{B,t}^d - V_t^*$. We apply this logic to different regions that span all possible values of V_t^*.

Region 1: $V_t^* \leq \lambda M_L^2$. There is no communication between the two parties and $\Pi_{B,t}^{c,PC}$ is always zero.

Region 2: $\lambda M_L^2 \leq V_t^* \leq \left(\frac{1+2\sqrt{2}}{2}\right)\lambda M_L^2$. In this region, the bank cannot commit not to trade starting at $\Delta X_{N,t} = 4\sqrt{V_t^*/\lambda} \geq 4M_L$. Meanwhile, since $\delta x_t^* \leq (4+2\sqrt{2})M_L$ so the bank's profit given that $\Delta X_{N,t}$ is between $4\sqrt{V_t^*/\lambda}$ and δx_t^* is $\frac{\lambda}{16}\Delta X_{N,t}^2 - V_t^*$. Otherwise it is 0.

Region 3: $\left(\frac{1+2\sqrt{2}}{2}\right)\lambda M_L^2 \leq V_t^* \leq \left(\frac{3+2\sqrt{2}}{2}\right)\lambda M_L^2$. This is similar to region 2, with the exception that $\delta x_t^* \geq (4+2\sqrt{2})M_L$, so the bank's profit given that $\Delta X_{N,t}$ is between $(4+2\sqrt{2})M_L$ and δx_t^* is $\frac{\lambda}{8}\Delta X_{N,t}^2 - \frac{\lambda}{2}M_L\Delta X_{N,t} + \frac{\lambda}{2}M_L^2 - V_t^*$.

Region 4: $V_t^* \geq \left(\frac{3+2\sqrt{2}}{2}\right)\lambda M_L^2$. Similar to region 3 but the point where the bank starts trading is now above $(4+2\sqrt{2})M_L$ and equals $2M_L + 2\sqrt{2M_L^2 + \frac{2V_t^*}{\lambda}}$. Therefore, whenever the bank makes positive profits, it makes $\frac{\lambda}{8}\Delta X_{N,t}^2 - \frac{\lambda}{2}M_L\Delta X_{N,t} + \frac{\lambda}{2}M_L^2 - V_t^*$. Q.E.D.

Lemma 12: *The bank's cooperation profits, as a function of V_t^*, are given by:*

Region 1: $V_t^* \leq \lambda M_L^2$.

$$\Pi_{B,t}^{c,PC} = 0 \tag{63}$$

Region 2: $\lambda M_L^2 \le V_t^* \le \left(\frac{1+2\sqrt{2}}{2}\right)\lambda M_L^2$**.**

$$\Pi_{B,t}^{c,PC} = \begin{cases} 0 & \text{if } \Delta X_{N,t} \le 4\sqrt{\frac{V_t^*}{\lambda}} \\ \frac{\lambda}{16}\Delta X_{N,t}^2 - V_t^* & \text{if } 4\sqrt{\frac{V_t^*}{\lambda}} \le \Delta X_{N,t} \le 8M_L - 4\sqrt{2M_L^2 - \frac{V_t^*}{\lambda}} \\ 0 & \text{if } \Delta X_{N,t} \ge 8M_L - 4\sqrt{2M_L^2 - \frac{V_t^*}{\lambda}} \end{cases}. \tag{64}$$

Region 3: $\left(\frac{1+2\sqrt{2}}{2}\right)\lambda M_L^2 \le V_t^* \le \left(\frac{3+2\sqrt{2}}{2}\right)\lambda M_L^2$**.**

$$\Pi_{B,t}^{c,PC} = \begin{cases} 0 & \text{if } \Delta X_{N,t} \le 4\sqrt{\frac{V_t^*}{\lambda}} \\ \frac{\lambda}{16}\Delta X_{N,t}^2 - V_t^* & \text{if } 4\sqrt{\frac{V_t^*}{\lambda}} \le \Delta X_{N,t} \le (4+2\sqrt{2})M_L \\ \frac{\lambda}{8}\Delta X_{N,t}^2 - \frac{\lambda}{2}M_L\Delta X_{N,t} + \frac{\lambda}{2}M_L^2 - V_t^* & \text{if } (4+2\sqrt{2})M_L \le \Delta X_{N,t} \le \\ & 3M_L + \frac{2V_t^*}{\lambda}\left(\frac{1}{M_L}\right) \\ 0 & \text{if } \Delta X_{N,t} \ge 3M_L + \frac{2V_t^*}{\lambda}\left(\frac{1}{M_L}\right) \end{cases}. \tag{65}$$

Region 4: $V_t^* \ge \left(\frac{3+2\sqrt{2}}{2}\right)\lambda M_L^2$**.**

$$\Pi_{B,t}^{c,PC} = \begin{cases} 0 & \text{if } \Delta X_{N,t} \le 2M_L + 2\sqrt{2M_L^2 + \frac{2V_t^*}{\lambda}} \\ \frac{\lambda}{8}\Delta X_{N,t}^2 - \frac{\lambda}{2}M_L\Delta X_{N,t} + \frac{\lambda}{2}M_L^2 - V_t^* & \text{if } 2M_L + 2\sqrt{2M_L^2 + \frac{2V_t^*}{\lambda}} \le \\ & \Delta X_{N,t} \le 3M_L + \frac{2V_t^*}{\lambda}\left(\frac{1}{M_L}\right) \\ 0 & \text{if } \Delta X_{N,t} \ge 3M_L + \frac{2V_t^*}{\lambda}\left(\frac{1}{M_L}\right) \end{cases}. \tag{66}$$

Proof of Lemma 12: The proof follows immediately from Corollary 4 given the fact that: (i) no relationship is sustained if $V_t^* \le \lambda M_L^2$, (ii) the bank starts front-running and the relationship breaks down prior to $\Delta X_{N,t} = (4+2\sqrt{2})M_L$ if $\lambda M_L^2 \le V_t^* \le (1+2\sqrt{2})\lambda M_L^2/2$, (iii) the bank starts front-running prior to $(4+2\sqrt{2})M_L$ and the relationship breaks down after $(4+2\sqrt{2})M_L$ when $(1+2\sqrt{2})\lambda M_L^2/2 \le V_t^* \le (3+2\sqrt{2})\lambda M_L^2/2$, and (iv) the bank starts front-running and the relationship breaks down after $(4+2\sqrt{2})M_L$ when $V_t^* \ge (3+2\sqrt{2})\lambda M_L^2/2$. Q.E.D.

Proof of Proposition 13: We proceed in a similar fashion to the proof of Proposition 10. Assume that $V_M^* < V_{PC}^*$ and that profits in both cases are given by $(\Pi_{A,M}^*, \Pi_{B,M}^*)$ and $(\Pi_{A,PC}^*, \Pi_{B,PC}^*)$. Consider the relational contract with:

$$\Pi^{**}_{B,M} = \begin{cases} \Pi^*_{B,M} & \text{if } \Delta X_{N,t} \leq \delta x^*_M \\ \\ \Pi^*_{B,PC} & \text{if } \Delta X_{N,t} > \delta x^*_{PC} \end{cases} . \tag{67}$$

As in Proposition 10, this profit function assures that both the arbitrageur's IR and the bank's IC constraints are met (the latter strictly). Therefore, there exists an $\epsilon > 0$ such that the relational contract with bank profits given by:

$$\Pi^{***}_{B,M} = \begin{cases} \Pi^*_{B,M} & \text{if } \Delta X_{N,t} \leq \delta x^*_M \\ \\ \Pi_M - \overline{\Pi}_A & \text{if } \delta x^*_M \leq \Delta X_{N,t} \leq \delta x^*_{PC} + \epsilon \\ \\ 0 & \text{if } \Delta X_{N,t} \geq \delta x^*_{PC} + \epsilon \end{cases} \tag{68}$$

that is self-enforcing and satisfies the surplus allocation rule in the monopoly setting. Since the breakdown in communication in this equilibrium is $\delta x^*_{PC} + \epsilon$, we have contradicted the statement that the initial monopoly equilibrium was efficient. Q.E.D.

Proof of Lemma 14: We explicitly prove the claim for the monopoly case where $V^*_M > (1 + 2\sqrt{2})\lambda M_L^2/2$. The relationship value in this case is determined by the fixed point problem:

$$V = \frac{\delta}{1-\delta} \int_{\frac{\beta\sigma}{4}}^{\frac{3\beta\sigma}{4} + \frac{8}{\beta\sigma}\left(\frac{V}{\lambda}\right)} \left[\frac{\lambda}{8}x^2 - \frac{\lambda\beta\sigma}{4}x + \frac{\lambda\beta 2\sigma^2}{8}\right] \left(\frac{1}{\sigma}\right) \sqrt{\frac{2}{\pi}} \exp\left[\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right] dx. \tag{69}$$

We can divide by $\lambda\sigma 2$ on both sides of this equation and rewrite it as:

$$\hat{V} = \frac{\delta}{1-\delta} \int_{\frac{\beta}{4}}^{\frac{3\beta}{4} + \frac{8\hat{V}}{\beta}} \left[\frac{1}{8}\hat{x}^2 - \frac{\beta}{4}\hat{x} + \frac{\beta 2}{8}\right] \sqrt{\frac{2}{\pi}} \exp\left[\frac{1}{2}\hat{x}^2\right] d\hat{x}. \tag{70}$$

where $\hat{x} = x/\sigma$ and $\hat{V} = V/\lambda\sigma^2$. It immediately follows that $\langle \delta x^* \rangle = 3\beta/4 + 8\hat{V}/\beta$ is a constant since $\hat{V}$ is the value of financial relationships when $\lambda = \sigma = 1$. The same procedure can be used to prove that $\langle \delta x^* \rangle$ is constant in the other cases since the bank's profit function is always quadratic in $\Delta X_{N,t}$. Q.E.D.

Remark on Lemma 14: The irrelevance of λ does not depend on the distribution of $\Delta X_{N,t}$. Likewise, the statement regarding σ will hold for any distribution whose pdf can be written in the form $f(x;\sigma) = g(x/\sigma)/\sigma$. Many additional distributions have this property (e.g. exponential distribution).

Proof of Proposition 15: Assume the bank decides to set her employment offer to attract an arbitrageur of type $\Delta X_{N,t} = x$. She will set his (targeted) wage offer at $W = \overline{\Pi}_A(x)$ since she has monopoly power and minimizing W will also maximize her ability to screen out speculators. Translating into an expression of wage as a function of profit, by using $\Pi = \Pi_M = \lambda x^2/8$, yields $W(\Pi) = 2M_L\sqrt{2\lambda\Pi} - 2\lambda M_L^2$. The constraint that the wage function be non-decreasing implies that $W(\Pi) = \sup_{\hat{\Pi}<\Pi} W(\hat{\Pi})$ for all $\Pi \notin \Omega$.

We now proceed to prove the remainder of the proposition in four parts: (i) We show that arbitrators with opportunity $\Delta X_{N,t} \leq 4M_L$ are not hired by the bank, (ii) that $\overline{M}$ is bounded, (iii) that $\overline{w} = \max_{x\leq\overline{M}} E^G[W(X \cdot \tilde{\Pi})]$, and (iv) that the wage offer is capped at $\overline{W} = 4\lambda M_L\overline{M} - 2\lambda M_L^2$.

First Part: The bank chooses to screen out arbitrators of type $\Delta X_{N,t} \leq 4M_L$ because they are unconstrained and their outside option satisfies $\overline{\Pi}_A = \Pi_M$. As a result, hiring these types would not provide additional profits to the bank and would also act to tighten the screening condition for the speculators (since they would benefit from the wage component that attracts the these arbitrators. Therefore, $W(\Pi) = 0$ for $\Pi \leq 2\lambda M_L^2$.

Second Part: Assume that $\overline{M}$ is unbounded. If $W(.)$ is unbounded as well, then it is impossible to screen speculators. Meanwhile, if $W(.)$ is bounded, arbitrators with opportunities $\Delta X_{N,t}$ greater than a threshold, $\hat{\delta x}$, will choose not reject the employment offer since $\overline{\Pi}_A$ is unbounded. This implies that no arbitrageur employed by the bank will ever choose to trade more than $\hat{\delta x}/4$ units. Hence, it is unnecessary to leave $\overline{M}$ unbounded. Contradiction.

Third Part: If $\overline{w} < \max_{x\leq\overline{M}} E^G[W(X \cdot \tilde{\Pi})]$, then speculators are not screened and the bank is better off not hiring anyone. Meanwhile, if $\overline{w} > \max_{x\leq\overline{M}} E^G[W(X \cdot \tilde{\Pi})]$, the bank can hire additional arbitrators, with types just beyond $\hat{\delta x}$, at a profit without violating the speculators screening condition.[14]

Fourth Part: The arbitrageur's outside option at $\Delta X_{N,t} = \overline{M}/4$ is equal to $4\lambda M_L\overline{M} - 2\lambda M_L^2$. As a result, if $\overline{W} > 4\lambda M_L\overline{M} - 2\lambda M_L^2$, arbitrageurs of type greater than $\overline{M}/4$ will choose to be employed by the bank. However, since the bank screens out speculators, these arbitrageurs could renegotiate their employment agreement with the bank after they join the firm. The bank would be willing to renegotiate because only arbitrageurs with $\Delta X_{N,t} > \overline{M}/4$ would attempt to do so. Understanding this, speculators would recognize that their screening condition fails if they attempt to renegotiate and would choose to accept the bank's employment offer. In order for the contract to be renegotiation proof, the marginal arbitrageur who accepts to work for the bank must be of type $\Delta X_{N,t} = \overline{M}/4$ which implies that $\overline{W} = 4\lambda M_L\overline{M} - 2\lambda M_L^2$. Q.E.D.

Remark on Proposition 15: Not all arbitrageurs of type above $M_L/4$ and below $\overline{M}/4$ will choose to work for the bank.

[14] If the cumulative distribution function of F were not strictly monotonic, we wouldn't be able to rule out the possibility that $\overline{w} > \max_{x\leq\overline{M}} E^G[W(X \cdot \tilde{\Pi})]$.

Proof of Proposition 16: If $\delta_M^* > 4\overline{M}$, explicit contracting alone does not lead to more effective arbitrage activity than the purely relational environment. Furthermore, since the bank keeps her profits from explicit contracting when she deviates on implicit promises, her IC constraint becomes:

$$\Pi_{B,t}^c + V^* \geq \Pi_{B,t}^d + V_E^* \tag{71}$$

where $V_E^* > 0$ is the bank's discounted future profits from explicit contracting with arbitrageurs. In this case, the bank's $\underline{X}_{B,t}$ is strictly higher than before and breakdown occurs earlier for a given V^*. This implies that V^* is lower than in the relational environment and the limits to arbitrage problem is worsened. Q.E.D.

References

[1] Abolafia, M., 1996, Making Markets, *Harvard University Press.*

[2] Abreu, D., D. Pearce, and E. Stacchetti, 1990, Toward a Theory of Discounted Repeated Games with Imperfect Monitoring, *Econometrica* **58**, 1041-1063.

[3] Anton, J., and D. Yao, 1994, Expropriation and Inventions: Appropriable Rents in the Absence of Property Rights, *American Economic Review* **84**, 190-209.

[4] Anton, J., and D. Yao, 2002, The Sale of Ideas: Strategic Disclosure, Property Rights, and Contracting, *Review of Economic Studies* **69**, 513-531.

[5] Anton, J., and D. Yao, 2004, Little Patents and Big Secrets: Managing Intellectual Property, *Rand Journal of Economics* **35**, 1-22.

[6] Arrow, K., 1971, Essays in the Theory of Risk-Bearing, *Markham Publishing Company.*

[7] Baker, G., R. Gibbons, and K. Murphy, 1994, Subjective Performance Measures in Optimal Incentive Contracts, *Quarterly Journal of Economics* **109**, 1125-1156.

[8] Benabou, R., and G. Laroque, 1992, Using Privileged Information to Manipulate Markets: Insiders, Gurus, and Credibility, *Quarterly Journal of Economics* **107**, 921-958.

[9] Bernhardt, D., V. Dvoracek, E. Hughson, and I. Werner, 2004, Why Do Larger Orders Receive Discounts on the London Stock Exchange?, *Review of Financial Studies* **18**, 1343-1368.

[10] Brunnermeier, M., 2001, Asset Pricing Under Asymmetric Information, *Oxford University Press.*

[11] Brunnermeier, M., and L. Pedersen, 2005, Predatory Trading, *Journal of Finance* **60**, 1825-1863.

[12] Cestone, G., and L. White, 2003, Anticompetitive Financial Contracting: The Design of Financial Claims, *Journal of Finance* **58**, 2109-2141.

[13] Chemmanur, T., and P. Fulghieri, 1994a, Investment Bank Reputation, Information Production, and Financia Intermediation, *Journal of Finance* **49**, 57-79.

[14] Chemmanur, T., and P. Fulghieri, 1994b, Reputation, Renegotiation, and the Choice Between Bank Loans and Publicly Traded Debt, *Review of Financial Studies* **7**, 475-506

[15] Degranges, G., and T. Foucault, 2002, Reputation-Based Pricing and Price Improvements, *Journal of Economics and Business* **57**, 493-527.

[16] DeLong, B., A. Shleifer, L. Summers, and R. Waldmann, 1990, Noise Trader Risk in Financial Markets, *Journal of Political Economy* **98**, 703-738.

[17] Fudenberg, D., D. Levine, and E. Maskin, 1994, The Folk Theorem with Imperfect Public Information, *Econometrica* **62**, 997-1039.

[18] Gromb, D., and D. Vayanos, 2002, Equilibrium and Welfare in Markets with Financially Constrained Arbitrageurs, *Journal of Financial Economics* **62**, 361-407.

[19] Grossman, S., and J. Stiglitz, 1980, On the Impossibility of Informationally Efficient Markets, *American Economic Review* **70**, 393-408.

[20] Grossman, S., and M. Miller, 1988, Liquidity and Market Structure, *Journal of Finance* **63**, 617-633.

[21] Gupta, U., 2000, Done Deals, *Harvard Business School Press.*

[22] Hellman, T., and E. Perotti, 2011, The Circulation of Ideas in Firms and Markets, *Management Science* **Forthcoming**.

[23] Ko, J., 2009, Leveraged Investor Disclosures and Concentrations of Risk, *Journal of Financial Markets* **12**, 368-390.

[24] Kyle, A., 1985, Continuous Auctions and Insider Trading, *Econometrica* **53**, 1315-1335.

[25] Levin, J., 2003, Relational Incentive Contracts, *American Economic Review* **93**, 835-857.

[26] Liu, J., and F. Longstaff, 2004, Losing Money on Arbitrage: Optimal Dynamic Portfolio Choice in Markets with Arbitrage Opportunities, *Review of Financial Studies* **17**, 611-641.

[27] Lobo, M., B. Carlin, and S. Viswanathan, 2007, Episodic Liquidity Crises: Cooperative and Predatory Trading, *Journal of Finance* **62**, 2235-2274.

[28] Lowenstein, R., 2000, When Genius Failed: The Rise and Fall of Long-Term Capital Management, *Random House, Inc.*

[29] Rajan, R., and L. Zingales, 2001, The Firm as a Dedicated Hierarchy: A Theory of the Origins and Growth of Firms, *Quarterly Journal of Economics* **121**, 805-851.

[30] Rotemberg, J., and G. Saloner, 1986, A Supergame-Theoretic Model of Price Wars During Booms, *American Economic Review* **76**, 390-407.

[31] Sharpe, S., 1990, Asymmetric Information, Bank Lending, and Implicit Contracts: A Stylized Model of Customer Relationships, *Journal of Finance* **45**, 1069-1087.

[32] Shleifer, A., and R. Vishny, 1997, The Limits to Arbitrage, *Journal of Finance* **52**, 35-55.

[33] Stein, J., 2005, Why Are Most Funds Open-End? Competition and the Limits to Arbitrage, *Quarterly Journal of Economics* **120**, 247-272.

[34] Xiong, W., 2001, Convergence Trading with Wealth Effects: An Amplification Mechanism in Financial Markets, *Journal of Financial Economics* **62**, 247-292.