

Financial Relationships and the Limits to Arbitrage*

Jiro E. Kondo[†]

Dimitris Papanikolaou[‡]

Abstract

We propose a model of limited arbitrage based on financial relationships. Financially constrained arbitrageurs may choose to seek additional financing from banks that have the technology to profit from the strategies themselves. A holdup problem arises because banks cannot commit to providing capital. To minimize competition, arbitrageurs will choose to stay constrained and underinvest in the arbitrage unless banks have sufficient reputational capital. This problem arises when mispricing is largest. More competition among financiers, higher arbitrageur wealth and allowing for explicit contracts can worsen the hold-up problem. When arbitrage is risky, financial relationships are more valuable, mitigating the problem.

Classification codes: G10, G14, G23

Keywords: limits to arbitrage, financial relationships

*We thank Thierry Foucault and two anonymous referees for helpful comments and suggestions. We are especially grateful to Dimitri Vayanos for his guidance and support during the early stages of the project. We also thank Kevin Chu, Jeremy Ko, Steve Ross, and the seminar participants at MIT and Waseda University for comments and discussions.

[†]McGill University

[‡]Kellogg School of Management and NBER

A number of irregularities in financial markets are often attributed to a “separation of brains and capital,”—specifically, the distinction between arbitrageurs possessing specialized knowledge but limited capital and other investors with a desire to save resources but a limited understanding of financial markets (Shleifer and Vishny, 1997). However, this view is somewhat at odds with the recent growth in the financial sector, in terms of both human capital and financial resources (Philippon, 2007; Philippon and Reshef, 2012). We propose a resolution to this apparent puzzle, which is based on the insight that arbitrage opportunities are ideas that cannot be legally protected.

At the heart of our model lies the following tradeoff. Arbitrageurs have access to arbitrage opportunities but limited funds. Financiers have access to unlimited capital but have the ability to implement the arbitrageurs’ ideas themselves. Arbitrageurs can obtain additional capital from financiers in exchange for revealing their ideas to them. The key friction in our setting is that, when accessing this source of finance, arbitrageurs run the risk of being expropriated by these banks. Since courts, like most investors, cannot properly evaluate the communication between arbitrageurs and banks, contracts are incomplete and banks cannot explicitly commit to providing any capital ex-post. We argue that the fact that certain financiers have the ability to understand arbitrageurs’ trading strategies also gives them access to similar, if not superior, financial technology to execute these strategies themselves. As a result, arbitrageurs may choose to remain financially constrained rather than reveal their private information and risk being expropriated.

In equilibrium, arbitrage is limited when it is most profitable. Specifically, a bank’s concern for its reputation provides some discipline against expropriating arbitrageurs. However, the value of financial relationships as a disciplinary mechanism is limited. The loss of future rents from interacting with arbitrageurs (the cost of expropriation) depends on the profitability of the average arbitrage opportunity. By contrast, the temptation to expropriate can be substantially higher if the current arbitrage opportunity is sufficiently valuable. Since the

best arbitrage ideas are at risk of expropriation, they will be kept secret by arbitrageurs. Consequently, relationships fail exactly when they are needed the most.

We formalize these insights in a stochastic repeated game between arbitrageurs and financiers. A sequence of short-lived arbitrageurs, each with knowledge of a different arbitrage opportunity, interact with infinitely lived banks. Arbitrageurs seek financiers to obtain capital and in exchange disclose their arbitrage ideas. If a bank expropriates an idea, future generations of arbitrageurs will punish this behavior by refusing to do business with that bank. Since there is a large number of banks, this punishment is credible. In equilibrium, banks can commit not to steal all but the most profitable ideas. The arbitrage ideas that get disclosed are implemented optimally, since banks and arbitrageurs collude to extract the full surplus. However, aware of the fact that their best ideas will be stolen, arbitrageurs prefer using other sources of capital to implement them and therefore choose to be financially constrained. Thus, the separation of brains and capital arises endogenously.

Our model delivers several novel predictions. We demonstrate that the limits-to-arbitrage problem becomes more substantial as competition increases between investment banks in attracting business from arbitrageurs. Increased competition among financiers lowers their share of surplus and therefore their incentives not to expropriate. Similarly, the ability of banks to commit decreases with arbitrageur wealth. Arbitrageurs' initial wealth determines their outside option relative to obtaining funding from banks; as this outside option increases, financial relationships become less valuable and therefore a less effective disciplining mechanism.

The key prediction of our model is that there is underinvestment in the most profitable arbitrage opportunities. This prediction is robust to allowing for explicit contracts or for risky arbitrage opportunities. Allowing for explicit contracts that are contingent on the payoff of the arbitrage strategy mitigates but does not eliminate this friction. Under the optimal contract, the arbitrageur's payoff is capped—to screen out low-quality arbitrageurs with incentives to

risk shift. This cap also effectively screens out high-quality arbitrageurs. Importantly, the introduction of explicit contracts also lowers the value of implicit relationships; hence, the net effect is ambiguous.

We find that increasing the riskiness of arbitrage opportunities can increase the value of financial relationships. For instance, banks can offer to “bail out” arbitrageurs in states when mispricing becomes worse. If arbitrageurs have nothing left to lose in these states, they will always accept the bank’s help. By contrast, if banks can make arbitrageurs worse off—for instance, by forcing them to liquidate their existing positions through predatory trading—then again the arbitrageurs with the best ideas will refuse help from the financiers, since they anticipate being expropriated. However, in both cases, since the bank can now capture additional rents, relationship values are higher, and therefore financiers can commit to not expropriating better ideas *ex-ante*.

Our paper complements the literature on limits to arbitrage. Existing models of limited arbitrage typically introduce a source of nonfundamental risk, implying that arbitrage opportunities are risky investments (e.g., De Long et al., 1990; Shleifer and Vishny, 1997; Xiong, 2001; Gromb and Vayanos, 2002; Kondor, 2009). In most of 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 arbitrageurs’ investment opportunities. 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. We maintain the first assumption, but relax the second. Absent the holdup problem we introduce, relaxing the second assumption eliminates limits to arbitrage.

Our paper contributes to this literature by providing a new model of the limits to arbitrage.

A key distinction with the existing literature is that, in our setting, arbitrage is limited due to the fear of expropriation by informable financiers, rather than the fear of trading losses. As a result, our model delivers sharply different predictions. In our setting, the value of relationships between arbitrageurs and financiers serves to ameliorate the holdup friction. Consequently, arbitrage is more limited when relationships are less valuable—specifically, when arbitrageurs have higher wealth or arbitrage opportunities are less risky. Last, our paper complements Attari et al. (2005), who explore the extent of limited arbitrage when financially constrained arbitrageurs interact with strategic traders who can predict their trades after initial positions are taken.

In addition to contributing to the limits-to-arbitrage literature, our paper is also related to several other strands of finance. Our work identifies a key instance in which some investors can take advantage of others through knowledge of their proprietary strategies and trading needs. Along similar lines, Brunnermeier and Pedersen (2005) study predatory trading in response to the predictable activity of distressed investors. La’O (2013) highlights how traders may forgo raising capital when this action signals distress and induces predatory trading by other traders in the market. Ueda (2004) considers the expropriation of ideas in the venture capital setting. Ko (2009) shows that arbitrageurs may be reluctant to disclose their positions to banks when the latter can front-run their trades and demonstrates that this can lead to endogenous concentrations of risk. Our work contributes to this literature by focusing on the limits to arbitrage and by explicitly considering financial relationships. Along these lines, our work is related to studies analyzing whether banks’ reputational concerns lead to improvements in efficiency (e.g., Sharpe, 1990; Chemmanur and Fulghieri, 1994b,a).¹

Our paper is part of a growing literature that explicitly incorporates financial relationships in the study of financial markets (see also Benabou and Laroque, 1992; Desgranges and

¹Our model is also similar to Rotemberg and 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.

Foucault, 2005; Bernhardt et al., 2005; Carlin et al., 2007). The impact of a lack of anonymity among financial market participants, especially large institutions, has remained largely unexplored. We contribute to this literature by studying how reputational concerns affect the financing and implementation of arbitrage activity. Closest to our paper is Carlin et al. (2007), who also use a stochastic repeated game to explore the role of relationships in determining whether large traders coordinate to provide liquidity or predate each other.

Last, our work 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 holdup problem, which is identical to ours, is known as the fundamental paradox (Arrow, 1976). There is a large literature that explores ways of mitigating this problem. Anton and 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 and Zingales (2001) study how organizational hierarchy can be used to minimize the problem of information leakage. Rather than looking at commitment, Anton and 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, Hellmann and 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.

1 The Model

Existing models of limited arbitrage feature arbitrageurs with ideas and limited capital. We extend this framework by introducing banks with unlimited capital and the ability to understand—and also expropriate—the arbitrageurs’ trading ideas. Hence, we typically refer to the bank as an *informable financier*. To illustrate the main mechanism in our model, we first consider a parsimonious framework where arbitrage opportunities are riskless and converge immediately following a round of trading. We relax this assumption in section 4.1.

An arbitrage opportunity consists of a pair of assets with identical payoffs but different prices. Specifically, 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, \bar{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. 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 nonstrategic 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. The degree of relative mispricing in the two markets is a function of the noise-trader spread, $\xi_t \equiv |X_{N1t} - X_{N2t}|$. We denote by F the cumulative distribution of ξ_t and assume that it has a finite second moment and is i.i.d. over time.

Long-term traders submit downward-sloping linear demands for each risky asset,

$$X_{LRi} = \frac{1}{\lambda} (\bar{v} - p_i), \quad (1)$$

where $\lambda > 0$. This specification of residual demand curves is standard in the literature (e.g., Brunnermeier and Pedersen, 2005; 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 and Miller, 1988) or uninformed investors fearing exploitation by informed ones (e.g., Grossman and Stiglitz, 1980).

We focus 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 .² Arbitrageurs live for one cycle and place convergence trades on A_1 and A_2 . A convergence trade is defined as a long position in A_i and an equally short position in A_j .

Arbitrageurs 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 and $M_L > 0$. The exact form of this constraint is not important; the key feature is that the arbitrageur has limited trading capacity. The portfolio constraint (2) captures the arbitrageurs' limited access to initial funding from other investors and banks that do not require her to explicitly disclose her trading ideas, for instance, when providing margin financing. Microeconomic foundations for this type of constraint are well established (see, e.g., Kiyotaki and Moore, 1997) and are commonly used in the literature (e.g., Gromb and Vayanos, 2002; Attari et al., 2005; Kondor, 2009; Grleanu and Pedersen, 2011).³

²Beyond 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.

³For example, this constraint can be motivated by credit-risk considerations. Suppose that financing is provided by banks and that there are two types of investors: risk-neutral speculators and arbitrageurs. Each investor has an initial wealth W and limited liability. However, unlike arbitrageurs, speculators have no information about asset payoffs that would allow for effective convergence trading. Instead, they invest in a random pair A_i and A_j whose payoffs are uncorrelated. Suppose that the value of each asset can change by an amount $[-\Delta, \Delta]$. If a speculator undertakes a convergence trade between two random assets to pool with arbitrageurs, he can lose up to 2Δ in wealth per unit of trade X_A . Since speculators have limited liability,

Arbitrageurs have the opportunity to relax their financial constraint (2) by interacting with financiers. We model the interaction between the arbitrageur and the bank as an extensive-form game. We refer to the stage game as an arbitrage opportunity cycle and summarize it in Figure 1. The arbitrageur identifies an opportunity at $t = 0$ and observes ξ , 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 $t = 3$. For now, all variables, including trading profits, are observable but not verifiable. We relax this assumption in Section 4.2.

The arbitrageur faces two choices when implementing the arbitrage opportunity. The first choice, denoted by $R = 0$, is to trade without revealing her strategy to the bank. In this case, the arbitrageur is subject to the default financial constraint. The bank does not trade in this case and prices are determined by market clearing:

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

Alternatively, the arbitrageur can choose to share her information to the bank in the hope of negotiating an increase in her trading capacity. This increase can be achieved with an infusion of capital or through a renegotiation of terms in the margin agreement. This choice is denoted by $R = 1$. The bank has the ability to understand the arbitrageur's strategy. This ability allows the bank to detect whether the arbitrageur misrepresents the proposed strategy; hence, the arbitrageur always has the incentive to tell the truth. However, this same ability also allows the bank to implement the trading strategy *without* the arbitrageur. Specifically,

if the bank wants to insure zero losses from lending to this type of investor, they would restrict X_A such that, $W - 2\Delta X_A \geq 0$, which is equivalent to setting $M_L = W(2\Delta)^{-1}$. If the proportion of investors who are speculators is high enough (e.g., investors can claim to have arbitrage opportunities), ruling out losses from them would be optimal for banks since speculators would trade as much as possible due to limited liability when $M_L > W/2\Delta$, and profits from arbitrageurs could not offset the bank's losses from this risk-shifting. In this setting, once the bank knows that the investor is an arbitrageur, it does not need to impose the constraint any more since the convergence trade will generate riskless profits.

the bank cannot commit ex-ante to alleviate the arbitrageur's financial constraint and may select any position limit satisfying $M \geq M_L$. The bank can also chose to compete with the arbitrageur by trading an amount $X_B \geq 0$ alongside him.

The arbitrageur chooses her own position, $X_A \leq M$, and all trades clear simultaneously at the market-clearing price:

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

Profits for the arbitrageur and bank are given by $\Pi_A \equiv \Delta p \cdot X_A$ and $\Pi_B \equiv \Delta p \cdot X_B$, respectively, where

$$\Delta p \equiv p_1 - p_2 = \lambda(\xi - 2X_A - 2X_B). \quad (5)$$

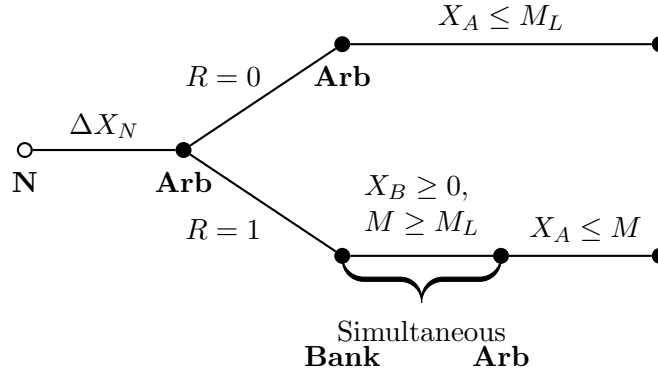


Figure 1: An arbitrage opportunity cycle

Last, we assume that arbitrageurs know the full history of the bank's behavior. This degree of knowledge could be rationalized if ξ_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. Allowing for imperfectly observable actions (as in Abreu et al. (1990) and Fudenberg et al. (1994)) would only amplify our limits to arbitrage problem. Since the bank will occasionally get away with misbehavior, and receive blame

when it behaves, its reputation loss following expropriation is lower.

2 Equilibrium

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 ξ_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 :*

$$V_t = E_t \left[\sum_{j=1}^{\infty} \delta^{t+j} \Pi_{B,t+j} \right]. \quad (6)$$

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 ξ_t is realized. 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).

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

2.1 Benchmark Equilibria

Prior to determining the equilibria of the game, it is instructive to look at three benchmark outcomes. The first benchmark 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 unconstrained arbitrageur

equilibrium (UA). The third benchmark corresponds to the case where the arbitrageur is financially constrained, and there is no financier that can relax her constraint. We refer to this as the no-bank equilibrium (NB). For brevity we relegate the details of these computations to the Appendix.

Benchmark NA: *In the no arbitrageur equilibrium, the price spread is given by*

$$\Delta p^{NA} = \lambda \xi. \quad (7)$$

Absent the arbitrageur, there exists a price spread among the two securities due to the presence of noise traders. The long-term traders absorb the excess supply ξ provided by noise traders and require an additional premium of λ per unit of demand imbalance.

Benchmark UA: *In the case where the arbitrageur is unconstrained, his demand is given by*

$$X_A^{UA} = \frac{1}{4} \xi, \quad (8)$$

the resulting price spread is

$$\Delta p^{UA} = \frac{\lambda}{2} \xi, \quad (9)$$

and arbitrageur profits are

$$\Pi^{UA} = \frac{\lambda}{8} \xi^2. \quad (10)$$

In this equilibrium, the arbitrageur's profits are maximized. The arbitrageur is the only agent possessing information about the mispricing and thus acts as a monopolist, taking into account her price impact. As a result, even when arbitrageurs are unconstrained, mispricing is not immediately eliminated. Here, we should note that this behavior is an artifact of the fact the model only has one period. In a dynamic model, arbitrageurs would trade continuously to eliminate the price spread as the number of trading periods goes to infinity.

Benchmark NB: *In the no-bank equilibrium, the arbitrageur's demand is given by*

$$X_A^{NB} = \min \left[M_L, \frac{1}{4} \xi \right], \quad (11)$$

the resulting price spread is

$$\Delta p^{NB} = \begin{cases} \frac{\lambda}{2} \xi & \text{if } \xi \leq 4M_L \\ \lambda (\xi - 2M_L) & \text{if } \xi \geq 4M_L \end{cases}, \quad (12)$$

and arbitrageur profits are

$$\Pi^{NB} = \begin{cases} \frac{\lambda}{8} \xi^2 & \text{if } \xi \leq 4M_L \\ \lambda M_L \xi - 2 \lambda M_L^2 & \text{if } \xi \geq 4M_L \end{cases}. \quad (13)$$

The no-bank equilibrium corresponds to a stylized version of the typical case considered in the literature. The arbitrageur has access to a riskless arbitrage opportunity, but limited capital. When the arbitrageur is constrained, equilibrium price spreads are above the unconstrained benchmark. Next, we analyze how arbitrageurs can potentially relax their portfolio constraint by revealing their arbitrage ideas in exchange for additional capital.

2.2 Static Game

To highlight the main agency friction in our model, we first focus on the case where financiers live for only one period. The following proposition characterizes the subgame perfect equilibria of the static game

Proposition 1 (Efficient SPE of the Static Game). *There is a unique SPE and the arbitrageur’s equilibrium strategy is*

$$R^* = 0 \quad \text{and} \quad X_A^* = \begin{cases} \frac{1}{4} \xi & \text{if } R = 0 \text{ and } \xi \leq 4M_L \\ \frac{1}{6} \xi & \text{if } R = 1 \text{ and } \xi \leq 6M_L \\ M_L & \text{otherwise} \end{cases} ,$$

while the bank’s equilibrium strategy is:

$$M^* = M_L \quad \text{and} \quad X_B^* = \begin{cases} 0 & \text{if } R = 0 \\ \frac{1}{6} \xi & \text{if } R = 1 \text{ and } \xi \leq 6M_L \\ \frac{1}{4} \xi - \frac{1}{2} M_L & \text{if } R = 1 \text{ and } \xi \geq 6M_L \end{cases} .$$

Since the bank cannot commit to relaxing the arbitrageur’s financial constraint or to not copying her strategy, it will always choose to expropriate the arbitrageur: the bank sets $M = M_L$ and $X_B > 0$ following communication. Because the bank has price impact, this unambiguously makes the arbitrageur worse off and she prefers not to reveal her information. As a result, arbitrage activity is limited in equilibrium.

Our agency friction provides a micro-foundation for the Shleifer and Vishny (1997) limited arbitrage. A key assumption in their setting is the “separation between brains and capital”—specifically, the absence of unconstrained investors with the ability to understand the arbitrageurs’ trading strategies. In our setting, these investors exist, but arbitrageurs are reluctant to access this source of funds because they are worried about expropriation. In equilibrium, both parties are worse off because of this agency problem. Banks could achieve positive profits while making the arbitrageur better off and markets more efficient if they could commit to loosening the arbitrageur’s financial constraint and limiting their proprietary trading.

The payoffs to the bank and the arbitrageur in the static game determine their outside

options in the repeated game. The bank's outside option in the repeated game equals its profits after the arbitrageur reveals her trading opportunity:

$$\Pi_B^d = \begin{cases} \frac{\lambda}{18} \xi^2 & \text{if } \xi \leq 6M_L \\ \frac{\lambda}{8} \xi^2 - \frac{\lambda}{2} M_L \xi + \frac{\lambda}{2} M_L^2 & \text{if } \xi \geq 6M_L \end{cases}. \quad (14)$$

The equilibrium payoff to the arbitrageur if she does not communicate her idea to the bank is given by equation (13) above,

$$\bar{\Pi}_{A,t} = \Pi^{NB} = \begin{cases} \frac{\lambda}{8} \xi^2 & \text{if } \xi \leq 4M_L \\ \lambda M_L \xi - 2 \lambda M_L^2 & \text{if } \xi \geq 4M_L \end{cases}. \quad (15)$$

The payoff in described in (15) determines the arbitrageur's outside option in the repeated game.

2.3 Repeated Game

Next, we analyze the repeated game between the financier and a sequence of short-lived arbitrageurs. We cast the repeated game in terms of a relational contract, that is, promises the bank makes to the arbitrageur regarding her behavior. In a given cycle, promises take the form $\Pi_{A,t}^c : \xi \rightarrow [0, \infty)$, specifying the profits offered to the arbitrageur as a function of ξ_t . Arbitrageurs cannot make any credible promises to the bank because they only live for one cycle.

The key object that determines equilibrium outcomes is the value of the relationship to the financier, which determines the cost of expropriating arbitrageurs. The value of this relationship equals the present value of the surplus that the financier acquires through its future interactions with arbitrageurs (6). The exact mechanism through which the surplus generated by the relationship is split between the arbitrageur and banks is not relevant to

our main conclusions. Hence, to simplify our discussion of the equilibrium, we allow for side payments between the arbitrageur and the bank. This assumption is innocuous since, in our setting, any profit-sharing agreement can be implemented through specific trading strategies by the arbitrageur and the bank in the states where cooperation occurs, $R = 1$. In the Appendix, we provide such an example. Analyzing situations in which the exact mechanism through which arbitrageurs and banks collude does affect equilibrium outcomes is an interesting extension of the model. We leave this for future research.

Efficient SPEs induce optimal collusion between the arbitrageur and the bank. In an efficient SPE, the arbitrageur and the bank optimally collude to maximize the total profits they can achieve:

$$\Pi_{A,t}^c + \Pi_{B,t}^c = \Pi_t^{UA}, \quad \text{if } R_t^* = 1. \quad (16)$$

When communication occurs, the total profits of the arbitrageur and bank equals that of the unconstrained equilibrium (10). Put simply, if the arbitrageur and the bank are going to collude, they will do so effectively. As a result, price spreads are as in the case where the arbitrageur is unconstrained.

In equilibrium, the arbitrageur will only choose to communicate her idea to the bank, $R_t^* = 1$, if the bank can credibly commit to satisfying her individual-rationality (IR) constraint:

$$\Pi_{A,t}^c \geq \bar{\Pi}_{A,t}. \quad (17)$$

Similarly, a bank's offer $\Pi_{A,t}^c$ is credible if and only if it is incentive compatible

$$\Pi_t^{UA} - \Pi_{A,t}^c + V_t^* \geq \Pi_{B,t}^d, \quad (18)$$

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 (IC) 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.

The level of $\Pi_{A,t}^c$ governs how the profits are split between the arbitrageur and the bank. In determining the properties of the equilibria of this model, the key question is whether these profits can be split in a way that satisfies both the arbitrageur's IR constraint and the bank's IC constraint. Using (16), we can rewrite the IR constraint as an upper bound on the bank's share of trading profits:

$$\Pi_{B,t}^c \leq \bar{\Pi}_{B,t} = \Pi_t^{UA} - \bar{\Pi}_{A,t} = \begin{cases} 0 & \text{if } \xi_t \leq 4M_L \\ \frac{\lambda}{8} \xi^2 - \lambda M_L \xi + 2\lambda M_L^2 & \text{if } \xi_t \geq 4M_L \end{cases}. \quad (19)$$

This bound is intuitive. If the arbitrageur is unconstrained (i.e., $\xi_t \leq 4M_L$), the bank cannot demand any profits without making him worse off. When the arbitrageur is constrained, there is some surplus to be shared between the two parties, and thus the bank can make positive profits.

Similarly, the bank's IC constraint implies a lower bound on the bank's equilibrium profits

$$\Pi_{B,t}^c \geq \underline{\Pi}_{B,t} = \Pi_B^d - V_t^* = \begin{cases} \max\left(\frac{\lambda}{18} \xi^2 - V_t^*, 0\right) & \text{if } \xi \leq 6M_L \\ \max\left(\frac{\lambda}{8} \xi^2 - \frac{\lambda}{2} M_L \xi + \frac{\lambda}{2} M_L^2 - V_t^*, 0\right) & \text{if } \xi \geq 6M_L \end{cases}. \quad (20)$$

This lower bound indicates that if the bank's temptation to deviate is high enough, she needs to obtain a certain level of profits or else she will choose to deviate.

We see that the IR and IC constraints can simultaneously hold if and only if

$$\Pi_t^{UA} - \bar{\Pi}_{A,t} \geq \Pi_{B,t}^d - V_t^*. \quad (21)$$

Combining (19) and (20), we see that both constraints can only hold if ξ is below a threshold

ξ^* , which is a function of the relationship value. As a result, communication between the arbitrageur and the bank are only possible if initial mispricing is small enough relative to the relationship value V_t^* . The following proposition formalizes this intuition.

Proposition 2 (Limits to Financial Relationships). *Information revelation by the arbitrageur cannot be sustained if $\xi_t > \xi_t^*$, where*

$$\xi_t^* = \begin{cases} 0 & \text{if } V_t^* \leq \frac{16}{18}\lambda M_L^2 \\ \frac{36}{5}M_L - \frac{12}{5}\sqrt{4M_L^2 - \frac{5}{2\lambda M_L}V_t^*} & \text{if } \frac{16}{18}\lambda M_L^2 \leq V_t^* \leq \frac{3}{2}\lambda M_L^2 \\ 3M_L + \frac{2}{\lambda M_L}V_t^* & \text{if } V_t^* \geq \frac{3}{2}\lambda M_L^2 \end{cases} . \quad (22)$$

Proposition 2 summarizes the first main result of the paper. Financial relationships are limited in their ability to mitigate the holdup problem between arbitrageurs and banks. The reason is that the bank's cost of expropriating the arbitrageur—its reputation value V_t^* —depends on the *average* quality of arbitrage opportunities. In contrast, the benefits to the bank from expropriation are increasing in ξ , as we see in equation (14). Consequently, the bank's incentive-compatibility constraint (18) cannot hold for all values of ξ . Furthermore, financial relationships are insufficient exactly when they are needed the most—when mispricing is large and arbitrages are most profitable.

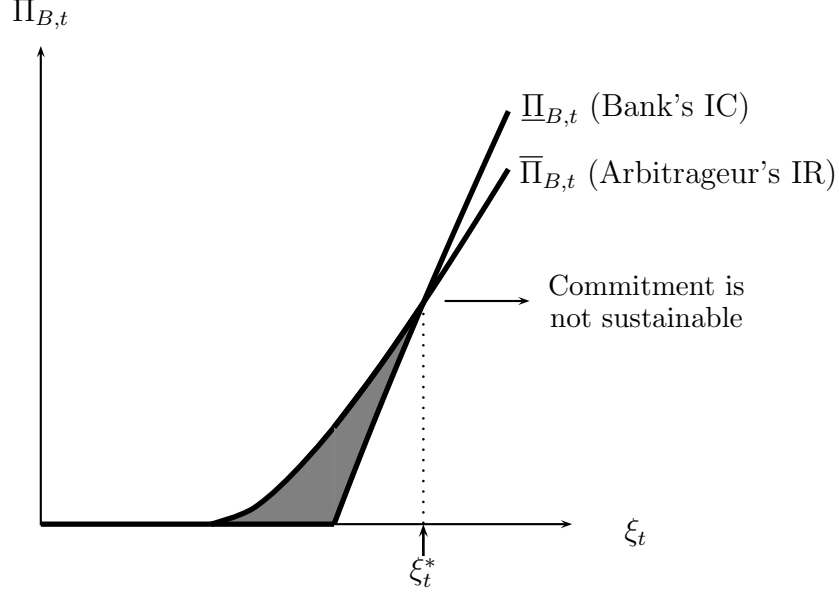


Figure 2: Illustration of Proposition 2. The figure plots the upper and lower bounds $[\underline{\Pi}_{B,t}, \bar{\Pi}_{B,t}]$ on the bank's cooperation profits as a function of ξ .

Figure 2 plots the upper and lower bounds $[\underline{\Pi}_{B,t}, \bar{\Pi}_{B,t}]$ on the bank's cooperation profits as a function of ξ . As we see, the arbitrageur's IR and the bank's IC constraint cannot both be satisfied for all values of mispricing ξ . All efficient equilibria of the dynamic game are indexed by the profit-sharing rule $\Pi_B^c(\xi)$. The next proposition completes our characterization of the efficient equilibria of the dynamic game.

Proposition 3 (Efficient SPE of the Dynamic Game). *In an efficient SPE*

- *The arbitrageur's communication strategy is given by*

$$R_t^*(\xi) = \begin{cases} 0 & \text{if } \xi \leq 4M_L \\ 1 & \text{if } \xi \in [4M_L, \xi_t^*] \\ 0 & \text{if } \xi \geq \xi_t^* \end{cases} . \quad (23)$$

- *Equilibrium price spreads are given by*

$$\Delta p_t^*(\xi) = \begin{cases} \frac{\lambda}{2} \xi_t & \text{if } \xi_t \leq \xi_t^* \\ \lambda \xi_t - 2\lambda M_L & \text{if } \xi_t \geq \xi_t^* \end{cases}. \quad (24)$$

- *Payoffs to the arbitrageur and the bank are*

$$\Pi_{A,t}^*(\xi) = \begin{cases} \frac{\lambda}{8} \xi^2 & \text{if } \xi_t \leq 4M_L \\ \frac{\lambda}{8} \xi^2 - \Pi_{B,t}^* & \text{if } \xi \in [4M_L, \xi_t^*] \\ \lambda M_L \xi - 2\lambda M_L^2 & \text{if } \xi_t \geq \xi_t^* \end{cases}, \quad (25)$$

and

$$\Pi_{B,t}^*(\xi) = \begin{cases} 0 & \text{if } \xi \leq 4M_L \\ \Pi_{B,t}^c & \text{if } \xi \in [4M_L, \xi_t^*] \\ 0 & \text{if } \xi \geq \xi_t^* \end{cases}. \quad (26)$$

where $\Pi_{B,t}^c$ satisfies equations (19)–(20).

The first part of Proposition 3 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.

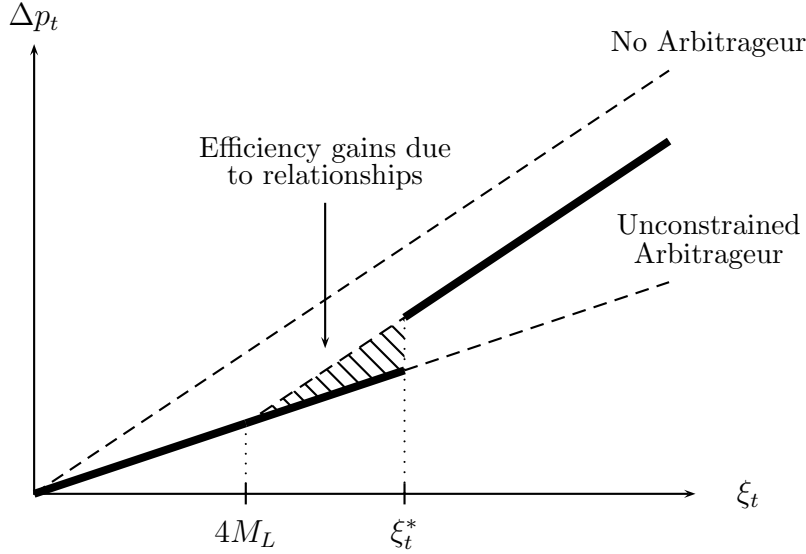


Figure 3: Equilibrium price spreads

Whenever there is communication between the arbitrageur and the bank, optimal collusion implies that price spreads are as in the unconstrained 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 $\xi_t = \xi_t^*$ versus $\xi_t = \xi_t^* + \varepsilon$, can lead to discontinuous changes in equilibrium mispricing. This is because relationships have their greatest effect in reducing mispricing at $\xi_t = \xi_t^*$ and no effect on price spreads once communication breaks down.

Last, we should emphasize that our main result cannot be obtained in a standard repeated game with ξ_t held constant across periods. To see this, we focus on the case where $\xi \geq 6M_L$ and the bank keeps all the surplus from its relationship with the arbitrageur. In the case

where mispricing is constant, the bank’s incentive compatibility constraint reduces to

$$\frac{\delta}{8} \left(\frac{\xi}{M_L} \right)^2 - \frac{1 + \delta}{2} \left(\frac{\xi}{M_L} \right) + \frac{3 + \delta}{2} \geq 0. \quad (27)$$

The key variable that determines the equilibrium is the degree of mispricing relative to the arbitrageur’s portfolio constraint M_L . Examining the bank’s IC constraint (27), we see that the opposite result obtains relative to the stochastic case: arbitrage is not limited when ξ is large relative to M_L . Specifically, the inequality (27) is guaranteed to hold for sufficiently high values of mispricing ξ/M_L . In contrast to the stochastic case, increasing ξ relative to M_L also increases the value of financial relationships. In particular, when ξ is high relative to M_L , the arbitrageur is more likely to be financially constrained, and therefore the surplus generated through her relationship with the financier is greater. Higher relationship values imply that the cost to the financier of expropriating arbitrageurs is greater, which mitigates the holdup problem and the resulting limits to arbitrage.

In sum, the main difference between this setting and ours is that allowing ξ 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. In this case, bank commitment becomes more difficult to sustain in times where ξ_t is high.

2.4 Arbitrageur Wealth and Limited Arbitrage

At the heart of our model lies a holdup friction. For this holdup problem to exist, the arbitrageur needs to have something to lose when communicating her idea to the bank in exchange for more capital. The higher her initial wealth, the more the arbitrageur stands to lose if she is expropriated. Under fairly general conditions, the limited arbitrage problem becomes worse in the infinitely repeated game as the arbitrageur’s initial wealth—which is proportional to M_L —increases. To simplify exposition, we first focus on the case where

the surplus allocation rule between the arbitrageur and the bank is independent of M_L . In line with the literature on Nash bargaining, surplus is given by the difference between the maximal profits Π^{UA} and the sum of the arbitrageur's outside option and the bank's net payoff from deviation:

$$S_t \equiv \Pi_{M,t} - [\bar{\Pi}_{A,t} + (\Pi_{B,t}^d - V_t)]. \quad (28)$$

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

$$\begin{aligned} \Pi_{B,t}^c &= (\Pi_{B,t}^d - V_t) + \alpha_t \cdot S_t \\ &= \alpha_t \cdot (\Pi_{M,t} - \bar{\Pi}_{A,t}) + (1 - \alpha_t) \cdot (\Pi_{B,t}^d - V_t). \end{aligned} \quad (29)$$

This result is illustrated in the following proposition

Proposition 4 (Arbitrageur Wealth and Limits to Arbitrage). *Let $\xi_t^*(X)$ denote the value of ξ_t^* in an equilibrium with $M_L = X$. Assuming that $0 < M_L' < M_L''$ and $\xi_t^*(M_L') > 0$, it follows that $\xi_t^*(M_L') > \xi_t^*(M_L'')$.*

Proposition 4 illustrates a key difference between our setting and the existing literature based on Shleifer and Vishny (1997). Increasing arbitrageur wealth leads to *more severe* limits to arbitrage. In equilibrium, increasing arbitrageur wealth has two effects that reduce the value of relationships to the bank V , and, as a result, the bank's ability to commit not to expropriate. 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. Both of these effects imply that the threshold ξ^* is lower, and the measure of arbitrage opportunities that are communicated to the bank falls. Since cooperation occurs in fewer states of the world, *both* parties are potentially worse off as M_L increases.

Importantly, arbitrage is limited in our setting precisely *because* $M_L > 0$. If the arbitrageur has no initial wealth and cannot raise *any* capital without communicating her strategy to the bank, $M_L = 0$, then her outside option is zero. In this case, the holdup friction disappears, since the arbitrageur is indifferent between being expropriated or not. In this case, the arbitrageur might as well communicate her idea for free, and therefore there are no limits to arbitrage. This extreme case best illustrates Proposition (4): our limits-to-arbitrage problem is least severe as the arbitrageur’s wealth goes to zero.

In our discussion so far, we have assumed that the arbitrageur’s bargaining power relative to the bank is independent of her initial wealth, M_L . Allowing for her bargaining power to increase with M_L leads to similar conclusions. As we see in the next section, increasing the surplus to the arbitrageur reduces the bank’s relationship value and thus worsens the limits-to-arbitrage problem.

3 Bank Competition and Limited Arbitrage

Our characterization of efficient equilibria in Proposition 3 allows for multiple equilibria, each indexed by the profit-sharing rule $[\Pi_{A,t}^c, \Pi_{B,t}^c]$. With the exception of Proposition (4), we have allowed $\Pi_{B,t}^c$ to be any curve in the shaded area of Figure 2—that is, anything between $\underline{\Pi}_{B,t}$ and $\bar{\Pi}_{B,t}$. To narrow the set of equilibria, we need specific assumptions about how surplus is split between the arbitrageur and the bank. 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 yields a unique equilibrium.

3.1 Simple Refinement

Here, we introduce competition among financiers in the repeated game. Specifically, we consider an extension of our stage game that allows banks to make offers of relational contracts

to the arbitrageur after she receives her private signal. Such promises prescribe a contingent action plan for the bank as a function of ξ_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.⁴ In the second case, there are a large number of potential financiers, leading to perfect competition in bidding. In both environments, stationarity ($V_t^* = V^*$) is driven by the independence assumption on ξ_t and the fixed surplus allocation rule.⁵

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 maximal level of profits (10) minus the arbitrageur's outside option (15). This implies a unique efficient SPE that is characterized by V_M^* . This relationship value is given by the largest solution to the fixed-point problem

$$V_M^* = \frac{\delta}{1 - \delta} \int_0^{\xi_M^*} \Pi_{B,t}^{c,M}(\xi) dF(\xi) \quad (30)$$

where the flow profits to the bank following communication are

$$\Pi_B^{c,M}(\xi) = \begin{cases} 0 & \text{if } \xi \leq 4M_L \\ \frac{\lambda}{8} \xi^2 - \lambda M_L \xi + 2\lambda M_L^2 & \text{if } 4M_L \leq \xi_t \leq \xi_M^* \\ 0 & \text{if } \xi \geq \xi_M^* \end{cases} \quad (31)$$

Existence of the unique solution, V_M^* , is guaranteed since zero is always a solution to the equation. 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

⁴Using the term monopoly here is a bit abusive since we have motivated some earlier results with the assumption that the arbitrageur can start up new relationships with other equally proficient banks.

⁵To obtain a stationary V^* , it is also necessary 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.

Pareto improvement. Therefore, only the equilibrium with V_t^{**} is an efficient SPE. 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 it possibly can. Hence, the bank either makes zero profits, or binds itself to its IC constraint,

$$\Pi_B^{c,PC}(\xi) = \max(\Pi_{B,t}^d - V_{PC}^*, 0), \quad (32)$$

where the continuation value V_{PC}^* is the largest solution to the problem,

$$V_{PC}^* = \frac{\delta}{1-\delta} \int_0^{\xi_{PC}^*} \Pi_B^{c,PC}(\xi) dF(\xi). \quad (33)$$

This produces a solution, V_{PC}^* , that is also independent of time. To conserve space, we relegate the equilibrium expression of $\Pi_B^{c,PC}$ to the Appendix.

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 banks' ability to make commitments. The following Proposition formalizes this intuition,

Proposition 5 (Monopoly versus Competition). *Comparing the efficient SPE from the monopoly and perfect competition cases, we have that:*

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

and

$$\xi_M^* \geq \xi_{PC}^*.$$

Proposition 5 illustrates another main result of the paper. When financiers compete away rents from financial relationships, the limits-to-arbitrage problem becomes worse. Broadly

interpreted, Proposition 5 has implications for the life cycle of the industry. When the industry is young and there are only a few specialized financiers, financial relationships have value and they serve to discipline bank behavior. As the industry matures and entry by new banks competes rents away, banks are tempted to behave more opportunistically as the cost of expropriating arbitrageurs falls. In this sense, we isolate a phenomenon that is broadly consistent with what Abolafia (2001) refers to as “cycles of opportunism” in financial markets.

3.2 Examples

Here, we work through two specific examples, one using a binomial and one using an exponential distribution.

3.2.1 Binomial Distribution

The simplest example that illustrates our mechanism is the binomial case, where ξ_t equals $\varepsilon > 0$ with probability p and 0 otherwise. Let $M_L = \beta\varepsilon/4$ with $\beta < 2/3$, so that the arbitrageur is constrained when there is mispricing and when the bank deviates. The smaller the value of β , the more severely constrained is the arbitrageur when $\xi_t = \varepsilon$.

When the bank has monopoly power, we can solve for the efficient equilibrium by assuming that information revelation can be sustained at $\xi_t = \varepsilon$ and checking for consistency afterwards. If there is communication, we have

$$V_M^* = \left(\frac{\delta}{1-\delta}\right) E[\Pi^{UA} - \bar{\Pi}_A] = \frac{\lambda p}{8} \left(\frac{\delta}{1-\delta}\right) (1-\beta)^2 \varepsilon^2. \quad (34)$$

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

$$\beta \leq \bar{\beta}_M = \frac{4\theta + 2 - 2\sqrt{\theta + 1}}{4\theta + 3}, \quad (35)$$

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

Notice that $\bar{\beta}_M$ is monotonically increasing in θ . This implies that there are no limits to arbitrage if the average level of mispricing is higher 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 $\bar{\beta}_M$. 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 values. 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 $\xi_t = \varepsilon$, we have

$$V_{PC}^* = \frac{\lambda\theta \left(\frac{1}{8} - \frac{\beta}{8} + \frac{\beta^2}{32} \right) \varepsilon^2}{1 + \theta}. \quad (36)$$

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

$$\beta \leq \bar{\beta}_{PC} = \frac{4\theta + 2 - 2\sqrt{\theta + 1}}{4\theta + 3}. \quad (37)$$

The same results and intuitions from the monopoly case hold here. Surprisingly, we have that $\bar{\beta}_{PC} = \bar{\beta}_M$ 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.

3.2.2 Exponential Distribution

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. Hence, we now consider an example where ξ is exponentially distributed. This example cannot be solved in closed form, so we provide numerical solutions.

We plot the threshold value ξ^* as a function of the surplus-sharing rule η between the arbitrageur and the bank and the arbitrageur's outside option M_L in Figure 4. Consistent with Proposition 5, we see that the relationship threshold ξ^* is increasing in the bank's share of the surplus η . Figure 4 illustrates that sometimes no relationship is feasible, that is, $\xi^* = 0$. Here, this happens when the bank's share of the surplus is sufficiently low or the arbitrageur's outside option is particularly high.

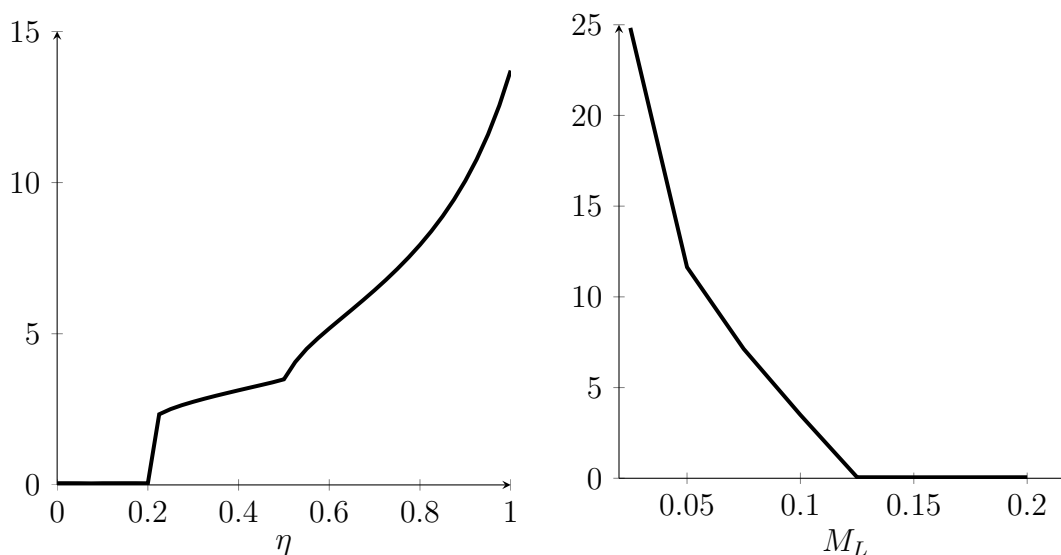


Figure 4: Figure plots the relationship threshold ξ^* as a function of the sharing rule η and the arbitrageur's financial constraint M_L . Unless otherwise noted, the figure uses $\delta = 0.8$, exponential distribution for ξ with mean equal to 2, $\eta = 0.5$, $\lambda = 1$, and $M_L = 0.1$.

Importantly, Figure 4 also indicates that a small change in parameter values can have a large and discontinuous impact on the role that relationships play in funding arbitrage

activity. This discontinuity arises because of the complementarity between ξ^* and V^* . For instance, an increase in the arbitrageur's outside option, M_L , implies that a lower share of total profits accrues to the bank, which lowers the value of relationships V^* . However, this reduction in value also lowers ξ^* , which further lowers V^* . As a result, small changes in parameters can have substantial effects on equilibrium outcomes.

4 Extensions

Here, we consider two extensions to the baseline model. First, we allow for risky arbitrage. Second, we allow for the possibility of explicit contracts.

4.1 Risky Arbitrage

We extend the model from Section 4 to allow for risky arbitrage. Specifically, we allow the mispricing in the two securities to get worse before it converges. This extension allows us to explore how our limits-to-arbitrage problem interacts with the traditional limits to arbitrage based on the dynamics of financial constraints (e.g., Shleifer and Vishny, 1997).

We introduce risk in the arbitrage opportunity by introducing an intermediate period t' within the arbitrage opportunity cycle, between t and $t + 1$. At time t' , the mispricing either worsens by an amount $q > 0$ with probability π or disappears:

$$\xi_{t'} = \begin{cases} \xi_t + q & \text{w/ probability } \pi \\ 0 & \text{otherwise} \end{cases} . \quad (38)$$

Next, it is necessary to specify how the default financial constraint changes at t' when mispricing worsens. We assume that the financial constraint tightens by an amount proportional

to the arbitrageur's losses between time t and t' ,

$$X_{A,t} + X_{A,t'} \leq M'_L = M_L - \theta q X_{A,t}, \quad (39)$$

where $X_{A,t'}$ is the arbitrageur's incremental trading at time t' and θ captures the sensitivity of the arbitrageur's financial constraint to his trading losses. The arbitrageur's financial constraint at time t is given by (2), as before. We focus on the case where $\theta q < 1$, so that if the arbitrageur invests up to her constraint at time t , $X_{A,t} = M_L$, the trading limit at time t' is positive, $M'_L > 0$.

This formulation captures the spirit of Shleifer and Vishny (1997). Specifically, arbitrageurs face a dynamic trade-off when engaging in arbitrage activity. If arbitrageurs trade more aggressively at time t , they may increase the likelihood that they are constrained in the future, t' if mispricing worsens. If that happens, they will be forced to liquidate their positions at a loss—and therefore sell when they should be buying. *Ceteris paribus*, arbitrageurs want to save some capital at time t , leading to underinvestment in the arbitrage opportunity.

As before, arbitrageurs have the option to reveal their arbitrage ideas to the bank in the hope of obtaining additional capital. However, the difference is that they can choose *when* to reveal their information. Specifically, we assume that the arbitrageur can reveal her idea to the bank either at t or t' . If the arbitrageur reveals her idea to the financier at time t , she risks being expropriated as before. However, if the arbitrageur reveals her idea at time t' , then she has the opportunity to build a constrained position in the arbitrage at time t . For now, we assume that if the bank expropriates the arbitrageur at time t' , then it cannot force her to liquidate any position she accumulated at time t .

To simplify exposition, we make the following simplifying assumptions. First, a different

group of long-term traders is active at times t and t' , implying that

$$\begin{aligned} \Delta p_t &= \lambda(\xi_t - X_t) \\ \Delta p_{t'} &= \begin{cases} \lambda(\xi_t + q - X_{t'}) & \text{w/ probability } \pi \\ 0 & \text{otherwise} \end{cases}, \end{aligned} \quad (40)$$

where $X_t = X_{A,t} + X_{B,t}$ is the combined trading by the arbitrageur and the bank at time t . This assumption ensures that the arbitrageur's only inter-temporal trade-off is due to the existence of the financial constraint (39). Second, we restrict attention to the case where the distribution of noise-trader demands satisfies $\xi \geq 4M_L - q$. This restriction ensures that the financial constraint is always binding at time t' . Third, we assume that the likelihood of the mispricing worsening is sufficiently small, that is, $\pi(1 + \theta q) < 1$. This parameter restriction ensures that the amount the arbitrageur trades at time t is increasing in ξ . Fourth, if the price-spread widens, the arbitrageur has zero bargaining power. Last, there is no discounting within an arbitrage opportunity cycle. None of these assumptions are crucial to our analysis; however, they do serve to simplify the exposition of the main ideas in this section.

At this stage, it is useful to describe the unconstrained benchmark. If the arbitrageur is not subject to any type of financial constraint, her optimal demands $X_{A,t}^{UA}$ and $X_{A,t'}^{UA}$ are given by

$$X_{A,t}^{UA} = \frac{\xi}{4} \quad \text{and} \quad \frac{\xi + q}{4}, \quad (41)$$

while her expected profits equal

$$\Pi_A^{UA}(\xi_t) = \frac{\lambda}{8} \left(\xi^2 + \pi(\xi + q)^2 \right). \quad (42)$$

Comparing the unconstrained total profits in the dynamic case (42) relative to the static case (10), we see that the possibility of worse mispricing at time t' allows for additional

profits for unconstrained arbitrageurs.

4.1.1 A Benchmark without Communication

First, we solve the model without communication between the arbitrageur and the bank. As before, the solution to this problem yields the arbitrageur's outside option in the full game. The arbitrageur solves the following optimization problem:

$$\bar{\Pi}_{A,t} = \max_{X_A} \lambda (\xi_t - X_{A,t}) X_{A,t} + \pi \lambda (\xi_t + q - X_{A,t'}) X_{A,t'}, \quad (43)$$

subject to the two portfolio constraints given by (2) and (39).

Recall that since $\pi(1 + \theta q) < 1$, the constraint (39) is always binding: if the mispricing widens, the arbitrageur needs to liquidate her position. As a result, the arbitrageur faces a dynamic trade-off between investing in the arbitrage opportunity at time t versus time t' . This trade-off can be clearly seen by examining the first-order condition of the problem in (43) and focusing on the case where the first period constraint (2) is not binding:

$$(\xi - 4 X_{A,t}) = \pi(1 + \theta q) \left(\xi + q(1 + 4\theta X_{A,t}) + 4 X_{A,t} - 4 M_L \right). \quad (44)$$

The arbitrageur equalizes the marginal benefit of trading an additional unit in the first period, to its marginal cost. The benefit is captured by the left-hand side in (44), which represents the impact of trading on profits in the first period t . The cost is represented by the right-hand side of (44), which captures the negative effect of trading at t on future profits at t' due to the possibility of the constraint being tighter in the future.

The arbitrageur's optimal trading strategy at time t equals

$$X_{A,t} = \min \left[M_L, \frac{1}{4} \frac{\xi - \pi(1 + \theta q)(\xi + q - 4 M_L)}{1 + \pi(1 + \theta q)^2} \right]. \quad (45)$$

Examining the arbitrageur’s investment strategy (45), we see that she trades less aggressively relative to the case where arbitrage is riskless (11). With probability π the mispricing may worsen at time t' , leading to trading losses for the arbitrageur and a tighter financial constraint $M_{L'}$. Under certain conditions—for instance, when mispricing ξ is large and the probability it increases π is small—the arbitrageur accepts the risk of possible liquidation at time t' , since $X_{A,t} > M_{L'}$.

In sum, the arbitrageur saves some of her capital at time t in case the mispricing worsens at time t' . Equation (45) captures the essence of Shleifer–Vishny’s limits to arbitrage within our setup. Since more aggressive trading at t forces the arbitrageur to take smaller positions at t' —when opportunities are more attractive—the arbitrageur views investment today as more costly due to noise trader risk and may limit the amount of arbitrage activity undertaken today.

4.1.2 Equilibrium with Communication

Next, we allow for communication between the arbitrageur and the bank. In contrast to the riskless setting, whether we allow for the possibility of cash transfers has a qualitative impact on our results. To simplify the analysis, we focus on the case where monetary transfers are allowed. This case represents the most efficient outcome; hence, the results in this section provide an upper bound on how efficient relationships can be in mitigating the limits-to-arbitrage problem in a risky setting.

The following proposition characterizes the equilibrium in the dynamic game with risky arbitrage.

Proposition 6 (Risky arbitrage). *In the equilibrium with risky arbitrage, there exists a finite threshold ξ^* such that (i) at time t , communication between the arbitrageur and the bank occurs for all values of $\xi_t \leq \xi^*$, (ii) when $\xi_t \leq \xi^*$ aggregate trading equals that of the unconstrained equilibrium at both t and t' and there are no limits to arbitrage, and (iii) when*

$\xi_t \geq \xi^*$, the arbitrageur always communicates his strategy at time t' if mispricing worsens regardless of the value of ξ .

Proposition (6) illustrates that arbitrage is still limited at time t when mispricing is large enough. However, whether the threshold ξ^* is lower or higher than in the riskless case depends on two offsetting forces.

First, the value of the relationship V to the bank is higher in the presence of risky arbitrage. This increase happens due to three effects. The first effect is that, due to the presence of the dynamic financial constraint, the arbitrageur's outside option (43) is lower when arbitrage is risky. Hence, the bank can appropriate more of the surplus. Second, the fact that mispricing can potentially worsen at time t' allows for more profit opportunities—for an unconstrained arbitrageur—as we see in equation (42). Thus, the total profits that can be shared between the arbitrageur and the bank are higher. The third effect is that the arbitrageur will always communicate her information to the bank if mispricing worsens at time t' . In contrast to time t , cooperation at time t' can be sustained for $\xi > \xi^*$, because the arbitrageur's outside option at time t' is *decreasing* in ξ . Specifically, if ξ is sufficiently high, the arbitrageur trades very aggressively at time t , accepting the fact that she needs to liquidate at time t' , $X_{A,t} > M_{L'}$. At this state, the arbitrageur is willing to reveal her information to the bank in exchange for a “bailout”, that is, a relaxation of her financial constraint. These three effects imply that the relationship value V and the surplus from the relationship (28) are unambiguously higher in this case relative to the riskless benchmark. As a result, holding the bank's deviation profits constant, cooperation can be sustained for higher values of ξ .

However, risky arbitrage implies the existence of an offsetting force. Specifically, the bank's profits from expropriating the arbitrageur at time t are higher when arbitrage is risky. In particular, arbitrageurs place more conservative demands, leaving more opportunities for the bank to expropriate the arbitrageur.

The net impact of these two opposing forces on the total surplus from relationships (28) is

ambiguous and depends on their relative magnitudes. In Figure 5 we compute the threshold ξ^* for different values of the probability of arbitrageur distress π and the sensitivity of financial constraint to trading losses θ . Overall, our numerical results suggest that risky arbitrage increases the value of relationships V and the communication threshold ξ^* .

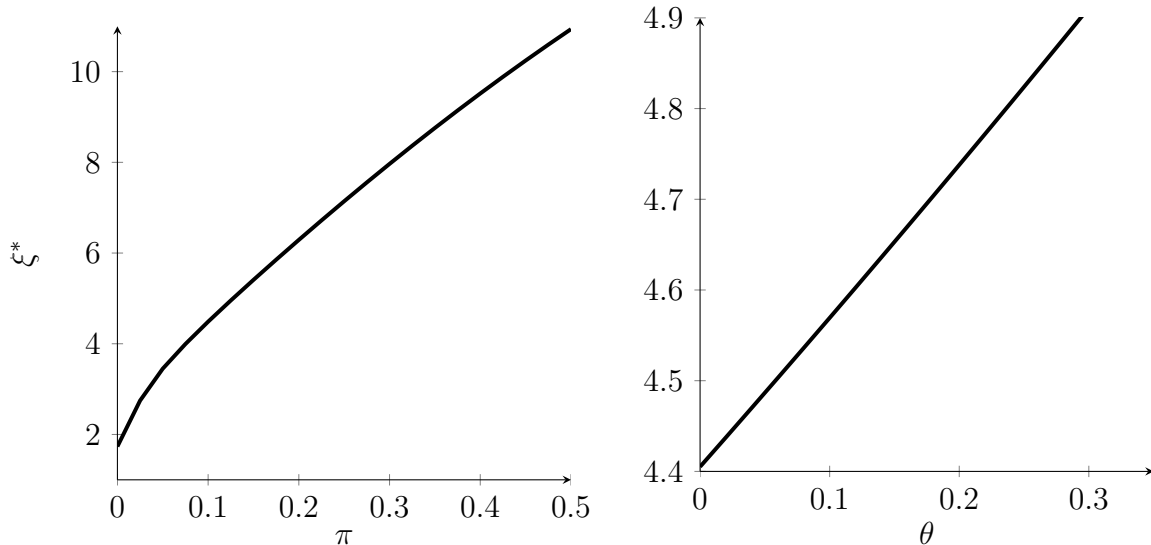


Figure 5: Sensitivity of relationship threshold ξ^* to key parameters. The figure plots the resulting relationship threshold ξ^* as a function of the probability of mispricing worsening, π and the sensitivity of the financial constraint to trading losses, θ . Unless otherwise noted, the figure uses $\delta = 0.9$, exponential distribution for ξ with mean equal to 2, $\eta = 0.5$, $\lambda = 1$, $\pi = 0.1$, $\theta = 0.05$, $q = 1$ and $M_L = 0.2$.

In this equilibrium, the arbitrageur always reveals her information to the bank when mispricing worsens at time t' . This behavior is consistent with anecdotal evidence, for example, the request for assistance from Goldman Sachs by LTCM during the Russian Crisis of 1998 (Lowenstein, 2000). However, the result that the arbitrageur is *always* willing to do so depends on the implicit assumption that she has nothing left to lose at time t' if mispricing gets worse. This assumption is not necessarily realistic. In many cases, financiers may be able to expropriate even more from distressed arbitrageurs forcing them to take even larger losses on their positions established at time t . If this is the case, then distressed arbitrageurs

may be reluctant to reveal their information at time t' .

To explore this possibility, we allow the bank to completely expropriate the arbitrageur. Specifically, we modify the existing setup to allow the bank to force the arbitrageur to liquidate her entire portfolio position, by setting $M_L = 0$ or $M_{L'} = 0$. In this case, the bank's deviation profits at time t' are *higher* than the profits in the unconstrained equilibrium, since the bank can force the arbitrageur to sell her existing holdings at time t . The following proposition characterizes the equilibrium in this case.

Proposition 7 (Risky arbitrage with complete expropriation). *In the equilibrium with risky arbitrage and complete expropriation, there exist two finite thresholds $\xi_1^* \leq \xi_2^*$ such that (i) at time t , communication between the arbitrageur and the bank occurs for all values of $\xi_t \leq \xi_1^*$, (ii) when $\xi_t \leq \xi_1^*$, aggregate trading equals that of the unconstrained equilibrium at both t and t' and there are no limits to arbitrage, (iii) when $\xi_t \in [\xi_1^*, \xi_2^*]$, the arbitrageur communicates her strategy at time t' if mispricing worsens and aggregate trading equals that of the unconstrained equilibrium, and (iv) when $\xi_t \geq \xi_2^*$ communication does not occur between the arbitrageur and the bank, and arbitrageur demands are given by (45).*

In sum, the results of this section illustrate that allowing for risky arbitrage does not alter the main qualitative predictions of our benchmark model. Arbitrage is still limited when it is needed the most. Further, allowing for risky arbitrage further highlights the differences in our setup from the standard limits to arbitrage formulation (Shleifer and Vishny, 1997). In contrast to the existing literature, the presence of risk *increases* the value of relationships and hence ameliorates the limits-to-arbitrage problem.

4.2 Explicit Contracting

In this section, we allow for the possibility of the arbitrageur signing explicit an labor contract with a bank. A labor contract specifies a nonnegative and nondecreasing wage payment that

is conditional on profits and possibly a trading budget that sets an upper bound on the position, \overline{M} , the arbitrageur (now a 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 its ability to attract employees even if it deviates from its implicit agreements with arbitrageurs who trade on their own.

Absent any other frictions, explicit contracts will overcome our limits-to-arbitrage problem since they allow the financier to ex-ante commit to a specific payoff to the arbitrageur as a function of ξ . However, this result is only true if all aspects of the problem are contractible. Indeed, the usefulness of relational contracts is that they allow the bank to screen *ex-post* certain characteristics that are not ex-ante contractible. To illustrate this idea, we introduce a new class of agents, termed “speculators”. From the perspective of the bank, speculators look identical to arbitrageurs but have no valuable arbitrage ideas. Instead, speculators take risky gambles at the bank’s expense.

As before, 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 ξ_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 1.⁶

Since speculation is costly, the bank will choose the explicit contract it offers to arbitrageurs

⁶This assumption is made for expositional reasons, and the results in this section continue to hold if $\theta \in (0, 1)$.

to screen out all speculators. Effectively, this introduces an additional “fly-by-night” constraint into the problem: speculators must prefer their outside option \bar{w} over their expected payoff under the optimal contract. The bank achieves this by introducing a position limit \bar{M} as part of the explicit contract that effectively caps the payoff.

The following proposition characterizes the equilibrium with explicit contracts.

Proposition 8 (Equilibrium with Explicit Contracts). *Let Ω denote the set of arbitrageur types ξ_t that are hired by the bank. We have that (i) the optimal employment contract specifies a finite upper bound \bar{M} , (ii) $\Omega \subseteq [M_L/4, \bar{M}/4]$, (iii) the speculators’ screening condition binds*

$$\bar{w} = \max_{X \leq \bar{M}} E^G[W(X \cdot \tilde{\Pi})], \quad (46)$$

(iv) the wage offer is given by

$$W(\Pi) = \begin{cases} 0 & \text{if } \Pi \leq 2\lambda M_L^2 \\ 2M_L\sqrt{2\lambda\Pi} - 2\lambda M_L^2 & \text{if } 2\lambda M_L^2 \leq \Pi \leq 2\lambda\bar{M}^2 \text{ and } 2\sqrt{\frac{2\Pi}{\lambda}} \in \Omega \\ \sup_{\hat{\Pi} < \Pi} W(\hat{\Pi}) & \text{if } 2\lambda M_L^2 \leq \Pi \leq 2\lambda\bar{M}^2 \text{ and } 2\sqrt{\frac{2\Pi}{\lambda}} \notin \Omega \\ 4\lambda M_L\bar{M} - 2\lambda M_L^2 & \text{if } \Pi \geq 2\lambda\bar{M}^2 \end{cases}, \quad (47)$$

and (v) Arbitrage is still limited.

The optimal contract features both a trading limit \bar{M} and an upper bound on wages \bar{W} . Both follow from i) the fly-by-night constraint, which induces the bank to limit the rents speculators can extract, and ii) 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.

Specifically, the bank needs to provide an upper bound on wages in order to screen out speculators. Given that wages are capped, no arbitrageur with opportunities ξ greater than a threshold, $\hat{\xi}$, where $\bar{\Pi}_A(\hat{\xi}) = \bar{W}$, will ever choose to take the employment offer, since their

outside option $\bar{\Pi}_A$ is unbounded. This implies that no arbitrageur employed by the bank will ever choose to trade more than $\hat{\xi}/4$ units. Hence, a contract that is renegotiation-proof will place a position limit \bar{M} .

Allowing for explicit contracts can actually *worsen* the limits to arbitrage friction. 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. Here, explicit contracts crowd out implicit contracts, as in Baker et al. (1994).

Proposition 9 (Crowding Out). *If $\xi_M^* > 4\bar{M} > 4M_L$, then allowing explicit contracts leads to more severe limits to arbitrage.*

We conclude that our main mechanism is robust to allowing for explicit contracts, once we recognize their limitations. Implicit contracts allow the bank to screen on nonverifiable information—here, the distinction between arbitrageurs and speculators. Last, the form of the explicit contract we obtain 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.

5 Empirical Implications

Our model has novel predictions that are unique to our setup. Specifically, our theory predicts that the degree of arbitrageur secrecy is positively correlated with the quality of arbitrage opportunities. Specifically, to the extent that there is a trade-off between secrecy and obtaining additional capital, our model predicts that arbitrageurs with the best ideas will be more secretive to avoid being expropriated by their financiers. Empirically identifying variations in arbitrageur secrecy can be challenging. However, one possibility is to consider the number of prime brokers employed by hedge funds. For instance, in our setting, the arbitrageur can decide to trade each leg of the arbitrage with a different institution. Consistent

with this view, anecdotal evidence suggests that hedge funds want to split their trades across multiple prime brokers to minimize the amount of learning by each prime broker about their strategies and preserve their secrecy (Lowenstein, 2000). Our theory thus would suggest a positive correlation between the number of prime broker used by hedge funds and the ex-post performance of these funds.

In addition, since our model complements the existing literature on limited arbitrage, it has implications for the relation of these theories to the data. Specifically, existing models of limited arbitrage have two main predictions: arbitrage activity is more limited when arbitrageurs are more wealthy and when arbitrage opportunities are riskier. However, these predictions are predicated on the assumption that arbitrageurs cannot collaborate with unconstrained providers of capital. By modeling this interaction, our theory not only proposes an additional mechanism through which arbitrage may be limited ex-ante, but also suggests that these two predictions may be more nuanced.

Increasing the wealth of arbitrageurs need not always improve the limits to arbitrage friction, since it affects how arbitrageurs interact with their financiers. In particular, our proposition 4 implies that increasing arbitrageur wealth, M_L , may in fact worsen the limits-to-arbitrage problem by reducing the value of financial relationships. Specifically, in our setting, the arbitrageur's initial wealth M_L determines not only her trading capacity when operating alone, but also the surplus she can extract from her relationship to the bank. Consequently, relaxing the arbitrageurs' portfolio constraint will reduce the efficiency of relationships and reduce the set of arbitrage opportunities that are optimally exploited (lower ξ^*). Whether the effect on reducing mispricing is positive or negative will depend on the size of the initial mispricing. For very profitable opportunities, $\xi \geq \xi^*$, the effect of increasing M_L will be positive, since in this instances, the arbitrageur was operating alone. However, for arbitrage opportunities close to the threshold ξ^* , the effect can be negative, since for these strategies relationships may no longer be feasible.

Our mechanism also implies that the relation between the level of risk and the degree of limited arbitrage may be ambiguous. Indeed, this prediction has met with mixed empirical support (see, e.g., Brav et al., 2010). As we illustrate in Section 4.1, financial relationships can become more valuable when arbitrage opportunities are more risky. Consequently, arbitrageurs who operate in more risky environments may effectively have access to more capital through their relationships with informable financiers, even though these arbitrageurs may operate more cautiously. In sum, the level of noise-trader risk need not be unconditionally positively correlated with the degree of mispricing in the market.

6 Conclusion

We provide an economic foundation for the limits-to-arbitrage problem, based on a holdup problem between arbitrageurs and banks (or other informable financiers). Reputational concerns on the part of banks partially mitigate this problem. However, relationships between banks and arbitrageurs fail when mispricing in markets is largest. 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 of informable finance. Finally, allowing for explicit contracts and risky arbitrage does not alter our main conclusions. In fact, both of these features can worsen the extent of the limits-to-arbitrage problem by lowering the mispricing threshold at which cooperation between the arbitrageur and the bank fails.

Our work has implications about the structure of the market for funding arbitrageurs. 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 in which they invest. 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, and hence face some constraints in expropriating arbitrageurs. Our theory illustrates that there are benefits to setting up institutions that can serve as informable financiers without having a substantial ability to expropriate the arbitrageur.

Last, our mechanism is broadly applicable to a variety of settings. For instance Kondo and Papanikolaou (2013) build upon the mechanism in this paper to study frictions in the sale of innovation. In particular, they embed this mechanism in a fully dynamic general equilibrium model to study how frictions in the sale of ideas interact with news about the arrival of future technologies.

Analytical Appendix

This section contains the unproven propositions from the paper.

Proof of Proposition 1. 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^{NB} . Therefore,

$$X_A^* = \begin{cases} \frac{1}{4}\xi & \text{if } R = 0 \text{ and } \xi \leq 4M_L \\ M_L & \text{if } R = 0 \text{ and } \xi \geq 4M_L \end{cases}, \quad (\text{A.1})$$

and

$$\Delta p^* = \begin{cases} \frac{\lambda}{2}\xi & \text{if } R = 0 \text{ and } \xi \leq 4M_L \\ \lambda\xi - 2\lambda M_L & \text{if } R = 0 \text{ and } \xi \geq 4M_L \end{cases}. \quad (\text{A.2})$$

Payoffs are given by

$$\Pi_A^* = \begin{cases} \frac{\lambda}{8}\xi^2 & \text{if } R = 0 \text{ and } \xi \leq 4M_L \\ \lambda M_L \xi - 2\lambda M_L^2 & \text{if } R = 0 \text{ and } \xi \geq 4M_L \end{cases}. \quad (\text{A.3})$$

Now assume that $R = 1$. Given ξ and a forecast of X_B^* , the arbitrageur chooses his demand such that

$$\max_{X_A \leq M_L} \lambda X_A (\xi - 2X_A - 2X_B^*).$$

This implies that:

$$X_A^* = \begin{cases} \frac{1}{4}(\xi - 2X_B^*) & \text{if } R = 1 \text{ and } \xi \leq 4M_L + 2X_B^* \\ M_L & \text{if } R = 1 \text{ and } \xi \geq 4M_L + 2X_B^* \end{cases}. \quad (\text{A.4})$$

Meanwhile, given ξ and a forecast of X_A^* , the bank chooses her demand such that

$$\max_{X_B} \lambda X_B (\xi - 2X_B - 2X_A^*). \quad (\text{A.5})$$

This implies that:

$$X_B^* = \frac{1}{4}(\xi - 2X_A^*). \quad (\text{A.6})$$

As a result, there are two regions of ξ to consider: one where the arbitrageur is not constrained and another where the arbitrageur is constrained.

- *Region 1:* $\xi \leq 6M_L$. We have $X_A^* = X_B^* = \frac{1}{6}\xi$.
- *Region 2:* $\xi \geq 6M_L$. We have $X_A^* = M_L$ and $X_B^* = \frac{1}{4}\xi - \frac{1}{2}M_L$.

Therefore, we have:

$$X_B^* = \begin{cases} \frac{1}{6}\xi & \text{if } R = 1 \text{ and } \xi \leq 6M_L \\ \frac{1}{4}\xi - \frac{1}{2}M_L & \text{if } R = 1 \text{ and } \xi \geq 6M_L \end{cases} \quad (\text{A.7})$$

and

$$X_A^* = \begin{cases} \frac{1}{6}\xi & \text{if } R = 1 \text{ and } \xi \leq 6M_L \\ M_L & \text{if } R = 1 \text{ and } \xi \geq 6M_L \end{cases}. \quad (\text{A.8})$$

Meanwhile, price spreads and arbitrageur profits are given by

$$\Delta p^* = \begin{cases} \frac{\lambda}{3}\xi & \text{if } R = 1 \text{ and } \xi \leq 6M_L \\ \frac{\lambda}{2}\xi - \lambda M_L & \text{if } R = 1 \text{ and } \xi \geq 6M_L \end{cases} \quad (\text{A.9})$$

and

$$\Pi_A^* = \begin{cases} \frac{\lambda}{18}\xi^2 & \text{if } R = 1 \text{ and } \xi \leq 6M_L \\ \frac{\lambda}{2}M_L\xi - \lambda M_L^2 & \text{if } R = 1 \text{ and } \xi \geq 6M_L \end{cases} \quad (\text{A.10})$$

It is easy to verify that $\Pi_A^*(R=0) \geq \Pi_A^*(R=1)$ for all ξ and, hence, $R^* = 0$. ■

Corollary of Proposition 1: Using the relevant price spread and demand functions from Proposition 1, we can easily compute the deviation profits for the bank and the arbitrageur's outside option.

- *Bank's Deviation Profit:* If the arbitrageur does reveal his information to the bank, the bank's payoff is

$$\Pi_B^d \equiv \Pi_B^* = \begin{cases} \frac{\lambda}{32}\xi^2 & \text{if } R = 1 \text{ and } \xi \leq 6M_L \\ \frac{\lambda}{8}\xi^2 - \frac{\lambda}{2}M_L\xi + \frac{\lambda}{2}M_L^2 & \text{if } R = 1 \text{ and } \xi \geq 6M_L \end{cases}. \quad (\text{A.11})$$

- *Arbitrageur's Outside Option:* If the arbitrageur does not reveal his information to the bank, the arbitrageur's payoff is

$$\bar{\Pi}_A \equiv \Pi_A^* = \begin{cases} \frac{\lambda}{8} \xi^2 & \text{if } R = 1 \text{ and } \xi \leq 6M_L \\ \lambda M_L \xi - 2\lambda M_L^2 & \text{if } R = 1 \text{ and } \xi \geq 6M_L \end{cases}. \quad (\text{A.12})$$

■

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

$$\Pi_{At}^c + \Pi_{Bt}^c = \Pi_t^{UA} \quad (\text{A.13})$$

or, equivalently in terms of trade and financing,

$$X_{B,t}^* + M_t^* = \frac{1}{4} \xi_t \quad (\text{A.14})$$

and

$$X_{A,t}^* = M_t^* \quad (\text{A.15})$$

That is, the arbitrageur and the bank optimally collude to maximize their total profits.

Proof of Lemma 1. First, notice that if $X_{B,t} > 0$, then an unconstrained arbitrageur would choose $X_{A,t} > \frac{1}{4} \xi_t - X_{B,t}$ since his unwillingness to internalize the effect of his demands on the bank's profits would lead to total demands that exceed the unconstrained level. This implies that if the arbitrageur is constrained and $X_{B,t} + M_t = \frac{1}{4} \xi_t$, 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 $X_{B,t}^* + M_t^* \neq \frac{1}{4} \xi_t$ for some ξ_t . It follows that $\Pi_{A,t}^* + \Pi_{B,t}^* < \frac{\lambda}{8} \xi_t^2$. Consider an alternative set of strategies identical to this one with the exception that, at the aforementioned ξ_t , we have $X_{B,t}^{**} = \frac{2\Pi_{B,t}^*}{\lambda \xi_t}$ and $M_t^{**} = \frac{1}{4} \xi_t - X_{B,t}^{**}$. This implies that $\Pi_{A,t}^{**} + \Pi_{B,t}^{**} = \frac{\lambda}{8} \xi_t > \Pi_{A,t}^* + \Pi_{B,t}^*$ 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. ■

Proof of Proposition 2. We consider three different regions for relationship values: (i) $V_t^* \leq \frac{16}{18} \lambda M_L^2$, (ii) $\frac{16}{18} \lambda M_L^2 \leq V_t^* \leq \frac{3}{2} \lambda M_L^2$, and (iii) $V_t^* \geq \frac{3}{2} \lambda M_L^2$.

First Region: If $V_t^* \leq \frac{16}{18} \lambda M_L^2$, it is impossible to satisfy both the arbitrageur's IR and the bank's IC constraints at $\xi_t = 4M_L$. As a result, no communication can ever be sustained between the two and $\xi_t^* = 0$.

Second Region: In this case, the breakdown of communication happens before $\xi_t = 6M_L$ and ξ_t^* solves,

$$\frac{\lambda}{8} (\xi_t^*)^2 - \lambda M_L \xi_t^* + 2\lambda M_L^2 = \frac{\lambda}{18} (\xi_t^*)^2 - V_t^*, \quad (\text{A.16})$$

where the relevant root is the lower one (since the upper root is greater than $6M_L$ and the IC constraint is violated between the lower and upper roots).

Third Region: When $V_t^* \geq \frac{3}{2}\lambda M_L^2$, the breakdown occurs after $\xi_t = 6M_L$ and ξ_t^* solves:

$$\frac{\lambda}{8} (\xi_t^*)^2 - \lambda M_L \xi_t^* + 2\lambda M_L^2 = \frac{\lambda}{8} (\xi_t^*)^2 - \frac{\lambda}{2} M_L \xi_t^* + \frac{\lambda}{2} M_L^2 - V_t^*, \quad (\text{A.17})$$

which is simply a linear equation. ■

Proof of Proposition 3. We focus on proving the optimality of the arbitrageur's communication strategy since all other aspects of the proposition and equilibrium follow from computations undertaken in earlier proofs. Regarding the communication strategy, we split this proof into two components: (i) show that $R_t^* = 1$ for $\xi_t = \xi_t^*$, and (ii) show that if $R_t^* = 1$ for ξ_t and a different ξ_t' satisfies $4M_L \leq \xi_t' \leq \xi_t$, then $R_t^* = 1$ for ξ_t' . Both are proven by contradiction.

First Component: Suppose there exists a Pareto-optimal equilibrium where $R_t^* = 0$ for $\xi_t = \xi_t^*$. Consider a strategy profile that is identical to this equilibrium with the exception that it differs at the point $\xi_t = \xi_t^*$. At that point, set $X'_{B,t} = \frac{2(\Pi_{M,t} - \bar{\Pi}_{A,t})}{\lambda \xi_t^*}$ and $M'_t = \frac{2\bar{\Pi}_{A,t}}{\lambda \xi_t^*}$. It is easy to check that both the arbitrageur's IR and the bank's IC constraints are met for $\xi_t = \xi_t^*$ and for all ξ_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 ξ_t and $R_t^* = 0$ for ξ_t' . Since $R_t^* = 1$, we have $\bar{X}_t > \underline{X}_t$ for ξ_t . It follows that $\bar{X}'_t > \underline{X}'_t$ for ξ_t' as well. Pick any $X_B \in [\underline{X}'_t, \bar{X}'_t]$ and consider a strategy profile that is identical to the stated equilibrium except that it sets $R'_t = 1$ and $X_{B,t} = X_B$ for ξ_t' . It is easily verified that both the arbitrageur's IR and the bank's IC constraints are satisfied for ξ_t' and for all ξ_t where $R_t^* = 1$. This new strategy profile is subgame perfect and Pareto-dominates the aforementioned equilibrium. Contradiction. ■

Proof of Proposition 4. We proceed to prove this by contradiction. Assume there exists two equilibria with $M'_L < M''_L$ such that $\xi_t^*(M'_L) \leq \xi_t^*(M''_L)$. From Proposition 2, $\xi_t^*(M'_L) < \xi_t^*(M''_L)$ 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 $4M_L \leq \xi_t \leq \xi_t^*(M'_L)$. Now consider an alternative relational

contract for M'_L with:

$$\Pi_{B,t}^{c,**}(M'_L) = \begin{cases} \Pi_{B,t}^{c,*}(M'_L) & \text{if } \xi_t \leq \xi_t^*(M'_L) \\ \Pi_{B,t}^{c,*}(M''_L) & \text{if } \xi_t > \xi_t^*(M'_L) \end{cases}. \quad (\text{A.18})$$

It follows that $V_t^{**}(M'_L) > V_t^*(M''_L)$ and the bank's IC constraint is strictly satisfied up until $\xi_t^*(M''_L)$. Given optimal collusion, the arbitrageur's IR constraint is also met up until $\xi_t^*(M''_L)$ 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 $\varepsilon > 0$ such that we can construct a relational contract with,

$$\Pi_{B,t}^{c,***}(M'_L) = \begin{cases} \Pi_{B,t}^{c,*}(M'_L) & \text{if } \xi_t \leq \xi_t^*(M'_L) \\ \Pi_{B,t}^{c,*}(M''_L) & \text{if } \xi_t^*(M'_L) < \xi_t \leq \xi_t^*(M''_L) \\ \max_{\xi_t} \Pi_{B,t}^{c,*}(M''_L) & \text{if } \xi_t^*(M''_L) < \xi_t \leq \xi_t^*(M''_L) + \varepsilon \\ 0 & \text{if } \xi_t > \xi_t^*(M''_L) + \varepsilon \end{cases}. \quad (\text{A.19})$$

This contract satisfies $\xi_t^{***}(M'_L) > \xi_t^*(M''_L)$ and contradicts the statement that the equilibrium with $\xi_t^*(M'_L)$ is efficient. ■

Proof of Proposition 5. We proceed in a similar fashion to the proof of Proposition 4. 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:

$$\Pi_{B,M}^{**} = \begin{cases} \Pi_{B,M}^* & \text{if } \xi_t \leq \xi_M^* \\ \Pi_{B,PC}^* & \text{if } \xi_t > \xi_M^* \end{cases}. \quad (\text{A.20})$$

As in Proposition 4, 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 $\varepsilon > 0$ such that the relational contract with bank profits given by,

$$\Pi_{B,M}^{***} = \begin{cases} \Pi_{B,M}^* & \text{if } \xi_t \leq \xi_M^* \\ \Pi_M - \bar{\Pi}_A & \text{if } \xi_M^* \leq \xi_t \leq \xi_{PC}^* + \varepsilon \\ 0 & \text{if } \xi_t \geq \xi_{PC}^* + \varepsilon \end{cases} \quad (\text{A.21})$$

that is self-enforcing and satisfies the surplus allocation rule in the monopoly setting. Since the breakdown in communication in this equilibrium is $\xi_{PC}^* + \varepsilon$, we have contradicted the statement that the initial monopoly equilibrium was efficient. ■

Proof of Proposition 6. We proceed by backward induction. Consider the case where mispricing widens at time t' , and the arbitrageur has not revealed his strategy to the bank at time t . Let $X_{A,t}$ be the arbitrageur's position established at time t ; his financial constraint implies that

$$X_{A,t'} \leq M_L - (1 + \theta q) X_{A,t} \quad (\text{A.22})$$

Since $\xi \geq 4M_L - q$, the arbitrageur is always constrained at time t' even if he were to not trade at time t . Suppose for now that $X_{A,t} = M_L$. If he does not reveal his strategy to the bank, given our price formulation, his outside option is negative since he needs to liquidate his existing position. Since the arbitrageur has nothing to lose, he will always reveal his information to the bank t' . The bank agrees to bail out the arbitrageur—the arbitrageur is not forced to sell his position at time t' —leaving him with a payoff of zero at time t' . Hence, setting $X_{A,t} = M_L$ is optimal in this case. At time t' the bank makes profits equal to

$$\Pi_{B,t'} = \frac{1}{8} \lambda (q + \xi)^2. \quad (\text{A.23})$$

Since the bank makes additional profits in this state, the bank's relationship value is

$$V = E_t \left[\sum_{j=1}^{\infty} \delta^{t+j} \left(\int_0^{\xi^*} \Pi_{B,t+j} dF(\xi) + \pi \int_{\xi^*}^{\infty} \frac{1}{8} \lambda (q + \xi)^2 dF(\xi) \right) \right]. \quad (\text{A.24})$$

The remaining steps closely follow the proof of Proposition 2. To conserve space, we just sketch them out. Consider the bank's decision to expropriate the arbitrageur at time t . If the bank decides to expropriate the arbitrageur at time t , it loses its reputation value V , but gets to compete with the arbitrageur. The bank's temptation to deviate is increasing in ξ , whereas the cost is not. Hence the bank will expropriate the arbitrageur at time t for $\xi \geq \xi^*$. Last, note that this is not the most efficient equilibrium. The arbitrageur would be willing to reveal his strategy at time t' as long as the bank guarantees him a payoff greater than his outside option—which is negative. Yielding higher surplus to the bank at time t' increases V , which in turn increases ξ^* . The steps to construct this equilibrium closely follow the arguments above. ■

Proof of Proposition 7. As in the proof of Proposition 6, consider the case where mispricing widens at time t' , and the arbitrageur has not revealed her strategy to the bank at time t . In this case, the bank can set $M_{L,t'} = 0$, in order to force the arbitrageur to fully liquidate her position at time t' , making the arbitrageur strictly worse off. In this case, the bank's deviation profits at time

t' equal

$$\Pi_{B,t'}^d = \frac{1}{8} \lambda (q + \xi + 2X_{A,t})^2. \quad (\text{A.25})$$

Suppose as before that $X_A = M_L$. The total surplus from the relationship at time t' equals

$$S' = \frac{1}{2} \lambda (4\theta^2 q^2 - 1) M_L^2 + (q + \xi) \left(\theta q - \frac{1}{2} \right) \lambda M_L + V^*. \quad (\text{A.26})$$

Since $2\theta q < 1$, the surplus S_t is decreasing in ξ . Hence, as in the proof of Proposition 2 there can exist a threshold ξ_2^* , such that if $\xi < \xi_2^*$ the arbitrageur will not reveal her strategy at time t' . Now, the relationship value to the bank equals

$$V = E_t \left[\sum_{j=1}^{\infty} \delta^{t+j} \left(\int_0^{\xi_1^*} \Pi_{B,t+j} dF(\xi) + \pi \int_{\xi_1^*}^{\xi_2^*} \frac{1}{8} \lambda (q + \xi)^2 dF(\xi) \right) \right]. \quad (\text{A.27})$$

The remaining steps are as in the proof of Proposition 6. ■

Proof of Proposition 8. Assume that the bank decides to set the employment offer to attract an arbitrageur of type $\xi_t = x$. The bank will set the (targeted) wage offer at $W = \bar{\Pi}_A(x)$ since it has monopoly power and minimizing W will also maximize its 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 prove the remainder of the proposition in four parts. We show that (i) arbitrators with opportunity $\xi_t \leq 4M_L$ are not hired by the bank, (ii) \bar{M} is bounded, (iii) $\bar{w} = \max_{x \leq \bar{M}} E^G[W(X \cdot \tilde{\Pi})]$, and (iv) the wage offer is capped at $\bar{W} = 4\lambda M_L \bar{M} - 2\lambda M_L^2$.

First Part: The bank chooses to screen out arbitrators of type $\xi_t \leq 4M_L$ because they are unconstrained and their outside option satisfies $\bar{\Pi}_A = \Pi_M$. 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 these arbitrators). Therefore, $W(\Pi) = 0$ for $\Pi \leq 2\lambda M_L^2$.

Second Part: Assume that \bar{M} is unbounded. If $W(\cdot)$ is unbounded as well, then it is impossible to screen speculators. Meanwhile, if $W(\cdot)$ is bounded, arbitrators with opportunities ξ_t greater than a threshold, $\hat{\delta}x$, will choose to reject the employment offer since $\bar{\Pi}_A$ 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 \bar{M} unbounded. Contradiction.

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

would not be able to rule out the possibility that $\bar{w} > \max_{x \leq \bar{M}} E^G[W(X \cdot \tilde{\Pi})]$.

Fourth Part: The arbitrageur's outside option at $\xi_t = \bar{M}/4$ is equal to $4\lambda M_L \bar{M} - 2\lambda M_L^2$. As a result, if $\bar{W} > 4\lambda M_L \bar{M} - 2\lambda M_L^2$, arbitrageurs of type greater than $\bar{M}/4$ 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 $\xi_t > \bar{M}/4$ 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 $\xi_t = \bar{M}/4$ which implies that $\bar{W} = 4\lambda M_L \bar{M} - 2\lambda M_L^2$.

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

Proof of Proposition 9. If $\xi_M^* > 4\bar{M}$, explicit contracting alone does not lead to more effective arbitrage activity than the purely relational environment. Furthermore, since the bank keeps its profits from explicit contracting when it deviates on implicit promises, its IC constraint becomes,

$$\Pi_{B,t}^c + V^* \geq \Pi_{B,t}^d + V_E^* \quad (\text{A.28})$$

where $V_E^* > 0$ are the bank's discounted future profits from explicit contracting with arbitrageurs. In this case, the bank's $\underline{\Pi}_{B,t}$ 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. ■

References

- Abolafia, M., 2001, *Making Markets: Opportunism and Restraint on Wall Street* (Harvard University Press).
- Abreu, Dilip, David Pearce, and Ennio Stacchetti, 1990, Toward a theory of discounted repeated games with imperfect monitoring, *Econometrica* 58, pp. 1041–1063.
- Anton, James J, and Dennis A Yao, 1994, Expropriation and inventions: Appropriable rents in the absence of property rights, *American Economic Review* 84, 190–209.
- Anton, James J, and Dennis A Yao, 2002, The sale of ideas: Strategic disclosure, property rights, and contracting, *Review of Economic Studies* 69, 513–31.

- Anton, James J., and Dennis A. Yao, 2004, Little patents and big secrets: Managing intellectual property, *The RAND Journal of Economics* 35, pp. 1–22.
- Arrow, K.J., 1976, *Essays in the theory of risk-bearing* (North-Holland).
- Attari, Mukarram, Antonio S. Mello, and Martin E. Ruckes, 2005, Arbitraging arbitrageurs, *The Journal of Finance* 60, pp. 2471–2511.
- Baker, George, Robert Gibbons, and Kevin J Murphy, 1994, Subjective performance measures in optimal incentive contracts, *The Quarterly Journal of Economics* 109, 1125–56.
- Benabou, Roland, and Guy Laroque, 1992, Using privileged information to manipulate markets: Insiders, gurus, and credibility, *The Quarterly Journal of Economics* 107, 921–58.
- Bernhardt, Dan, Vladimir Dvoracek, Eric Hughson, and Ingrid M. Werner, 2005, Why do larger orders receive discounts on the london stock exchange?, *Review of Financial Studies* 18, 1343–1368.
- Brav, Alon, J.B. Heaton, and Si Li, 2010, The limits of the limits of arbitrage, *Review of Finance* 14, 157–187.
- Brunnermeier, Markus K., and Lasse Heje Pedersen, 2005, Predatory trading, *Journal of Finance* 60, 1825–1863.
- Brunnermeier, M.K., 2001, *Asset Pricing Under Asymmetric Information: Bubbles, Crashes, Technical Analysis, and Herding* (Oxford University Press).
- Carlin, Bruce Ian, Miguel Sousa Lobo, and S. Viswanathan, 2007, Episodic liquidity crises: Cooperative and predatory trading, *The Journal of Finance* 62, 2235–2274.
- Chemmanur, Thomas J, and Paolo Fulghieri, 1994a, Investment bank reputation, information production, and financial intermediation, *Journal of Finance* 49, 57–79.
- Chemmanur, Thomas J, and Paolo Fulghieri, 1994b, Reputation, renegotiation, and the choice between bank loans and publicly traded debt, *Review of Financial Studies* 7, 475–506.
- De Long, J Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann, 1990, Noise trader risk in financial markets, *Journal of Political Economy* 98, 703–38.
- Desgranges, Gabriel, and Thierry Foucault, 2005, Reputation-based pricing and price improvements, *Journal of Economics and Business* 57, 493 – 527.

- Fudenberg, Drew, David Levine, and Eric Maskin, 1994, The folk theorem with imperfect public information, *Econometrica* 62, pp. 997–1039.
- Grleanu, Nicolae, and Lasse Heje Pedersen, 2011, Margin-based Asset Pricing and Deviations from the Law of One Price, *Review of Financial Studies* 24, 1980–2022.
- Gromb, Denis, and Dimitri Vayanos, 2002, Equilibrium and welfare in markets with financially constrained arbitrageurs, *Journal of Financial Economics* 66, 361–407.
- Grossman, Sanford J., and Merton H. Miller, 1988, Liquidity and market structure, *The Journal of Finance* 43, pp. 617–633.
- Grossman, Sanford J., and Joseph E. Stiglitz, 1980, On the impossibility of informationally efficient markets, *The American Economic Review* 70, pp. 393–408.
- Hellmann, Thomas, and Enrico Perotti, 2011, The circulation of ideas in firms and markets, *Management Science* 57, 1813–1826.
- Kiyotaki, Nobuhiro, and John Moore, 1997, Credit cycles, *Journal of Political Economy* 105, 211–48.
- Ko, K. Jeremy, 2009, Leveraged investor disclosures and concentrations of risk, *Journal of Financial Markets* 12, 368 – 390.
- Kondo, Jiro E., and Dimitris Papanikolaou, 2013, News and cooperation cycles, Working paper, Northwestern University.
- Kondor, Peter, 2009, Risk in dynamic arbitrage: The price effects of convergence trading, *Journal of Finance* 64, 631–655.
- La’O, Jennifer, 2013, Predatory trading and credit freezes, Working paper, Columbia University.
- Levin, Jonathan, 2003, Relational incentive contracts, *American Economic Review* 93, 835–857.
- Lowenstein, Robert, 2000, *When Genius Failed: The Rise and Fall of Long-Term Capital Management* (Random House).
- Philippon, Thomas, 2007, Why has the U.S. financial sector grown so much? The role of corporate finance., Working Paper 13405, National Bureau of Economic Research.

- Philippon, Thomas, and Ariell Reshef, 2012, Wages and human capital in the U.S. finance industry: 1909–2006, *The Quarterly Journal of Economics* .
- Rajan, Raghuram G., and Luigi Zingales, 2001, The firm as a dedicated hierarchy: A theory of the origins and growth of firms, *The Quarterly Journal of Economics* 116, 805–851.
- Rotemberg, Julio J., and Garth Saloner, 1986, A supergame-theoretic model of price wars during booms, *The American Economic Review* 76, pp. 390–407.
- Sharpe, Steven A., 1990, Asymmetric information, bank lending and implicit contracts: A stylized model of customer relationships, *The Journal of Finance* 45, pp. 1069–1087.
- Shleifer, Andrei, and Robert W Vishny, 1997, The limits of arbitrage, *Journal of Finance* 52, 35–55.
- Ueda, Masako, 2004, Banks versus venture capital: Project evaluation, screening, and expropriation, *Journal of Finance* 59, 601–621.
- Xiong, Wei, 2001, Convergence trading with wealth effects: an amplification mechanism in financial markets, *Journal of Financial Economics* 62, 247–292.