THE JOURNAL OF FINANCE • VOL. LII, NO. 1 • MARCH 1997

The Limits of Arbitrage

ANDREI SHLEIFER and ROBERT W. VISHNY*

ABSTRACT

Textbook arbitrage in financial markets requires no capital and entails no risk. In reality, almost all arbitrage requires capital, and is typically risky. Moreover, professional arbitrage is conducted by a relatively small number of highly specialized investors using other people's capital. Such professional arbitrage has a number of interesting implications for security pricing, including the possibility that arbitrage becomes ineffective in extreme circumstances, when prices diverge far from fundamental values. The model also suggests where anomalies in financial markets are likely to appear, and why arbitrage fails to eliminate them.

ONE OF THE FUNDAMENTAL concepts in finance is arbitrage, defined as "the simultaneous purchase and sale of the same, or essentially similar, security in two different markets for advantageously different prices" (Sharpe and Alexander (1990)). Theoretically speaking, such arbitrage requires no capital and entails no risk. When an arbitrageur buys a cheaper security and sells a more expensive one, his net future cash flows are zero for sure, and he gets his profits up front. Arbitrage plays a critical role in the analysis of securities markets, because its effect is to bring prices to fundamental values and to keep markets efficient. For this reason, it is extremely important to understand how well this textbook description of arbitrage approximates reality. This article argues that the textbook description does not describe realistic arbitrage trades, and, moreover, the discrepancies become particularly important when arbitrageurs manage other people's money.

Even the simplest realistic arbitrages are more complex than the textbook definition suggests. Consider the simple case of two Bund futures contracts to deliver DM250,000 in face value of German bonds at time T, one traded in London on LIFFE and the other in Frankfurt on DTB. Suppose for the moment, counter factually, that these contracts are exactly the same. Suppose finally that at some point in time t the first contract sells for DM240,000 and the second for DM245,000. An arbitrageur in this situation would sell a futures contract in Frankfurt and buy one in London, recognizing that at time T he is perfectly hedged. To do so, at time t, he would have to put up some good faith money, namely DM3,000 in London and DM3,500 in Frankfurt, leading to a

* Shleifer is from Harvard University and Vishny is from The University of Chicago. Nancy Zimmerman and Gabe Sunshine have helped us to understand arbitrage. We thank Yacine Aït Sahalia, Douglas Diamond, Oliver Hart, Steve Kaplan, Raghu Rajan, Jésus Saa-Requejo, Luigi Zingales, Jeff Zwiebel, and especially Matthew Ellman, Gustavo Nombela, René Stulz, and an anonymous referee for helpful comments.

net cash outflow of DM6,500. However, he does not get the DM5,000 difference in contract prices at the time he puts on the trade. Suppose that prices of the two contracts both converge to DM242,500 just after t, as the market returns to efficiency. In this case, the arbitrageur would immediately collect DM2,500 from each exchange, which would simultaneously charge the counter parties for their losses. The arbitrageur can then close out his position and get back his good faith money as well. In this near textbook case, the arbitrageur required only DM6,500 of capital and collected his profits at some point in time between t and T.

Even in this simplest example, the arbitrageur need not be so lucky. Suppose that soon after t, the price of the futures contract in Frankfurt rises to DM250,000, thus moving further away from the price in London, which stays at DM240,000. At this point, the Frankfurt exchange must charge the arbitrageur DM5,000 to pay to his counter party. Even if eventually the prices of the two contracts converge and the arbitrageur makes money, in the short run he loses money and needs more capital. The model of capital-free arbitrage simply does not apply. If the arbitrageur has deep enough pockets to always access this capital, he still makes money with probability one. But if he does not, he may run out of money and have to liquidate his position at a loss.

In reality, the situation is more complicated since the two Bund contracts have somewhat different trading hours, settlement dates, and delivery terms. It may easily happen that the arbitrageur has to find the money to buy bonds so that he can deliver them in Frankfurt at time T. Moreover, if prices are moving rapidly, the value of bonds he delivers and the value of bonds delivered to him may differ, exposing the arbitrageur to additional risks of losses. Even this simplest trade then becomes a case of what is known as risk arbitrage. In risk arbitrage, an arbitrageur does not make money with probability one, and may need substantial amounts of capital to both execute his trades and cover his losses. Most real world arbitrage trades in bond and equity markets are examples of risk arbitrage in this sense. Unlike in the textbook model, such arbitrage is risky and requires capital.

One way around these concerns is to imagine a market with a very large number of tiny arbitrageurs, each taking an infinitesimal position against the mispricing in a variety of markets. Because their positions are so small, capital constraints are not binding and arbitrageurs are effectively risk neutral toward each trade. Their collective actions, however, drive prices toward fundamental values. This, essentially, is the model of arbitrage implicit in Fama's (1965) classic analysis of efficient markets and in models such as CAPM (Sharpe (1964)) and APT (Ross (1976)).

The trouble with this approach is that the millions of little traders are typically not the ones who have the knowledge and information to engage in arbitrage. More commonly, arbitrage is conducted by relatively few professional, highly specialized investors who combine their knowledge with resources of outside investors to take large positions. The fundamental feature of such arbitrage is that brains and resources are separated by an agency relationship. The money comes from wealthy individuals, banks, endowments, and other investors with only a limited knowledge of individual markets, and is invested by arbitrageurs with highly specialized knowledge of these markets. In this article, we examine such arbitrage and its effectiveness in achieving market efficiency.

In particular, the implications of the fact that arbitrage—whether it is ultimately risk-free or risky—generally requires capital become extremely important in the agency context. In models without agency problems, arbitrageurs are generally more aggressive when prices move further from fundamental values (see Grossman and Miller (1988), De Long et al. (1990), Campbell and Kyle (1993)). In our Bund example above, an arbitrageur would in general increase his positions if London and Frankfurt contract prices move further out of line, as long as he has the capital. When the arbitrageur manages other people's money, however, and these people do not know or understand exactly what he is doing, they will only observe him losing money when futures prices in London and Frankfurt diverge. They may therefore infer from this loss that the arbitrageur is not as competent as they previously thought, refuse to provide him with more capital, and even withdraw some of the capital—even though the expected return from the trade has increased.

We refer to the phenomenon of responsiveness of funds under management to past returns as performance based arbitrage. Unlike arbitrageurs using their own money, who allocate funds based on expected returns from trades, investors may rationally allocate money based on past returns of arbitrageurs. When arbitrage requires capital, arbitrageurs can become most constrained when they have the best opportunities, i.e., when the mispricing they have bet against gets even worse. Moreover, the fear of this scenario would make them more cautious when they put on their initial trades, and hence less effective in bringing about market efficiency. This article argues that this feature of arbitrage can significantly limit its effectiveness in achieving market efficiency.

We show that performance-based arbitrage is particularly ineffective in extreme circumstances, where prices are significantly out of line and arbitrageurs are fully invested. In these circumstances, arbitrageurs might bail out of the market when their participation is most needed. Performance based arbitrage, then, is even more limited than arbitrage described in earlier models of inefficient markets, such as Grossman and Miller (1988), De Long et al. (1990), and Campbell and Kyle (1993).

Ours is obviously not the first study of the consequences of delegated portfolio management. Early articles in this area include Allen (1990) and Bhattacharya-Pfleiderer (1985). Scharfstein and Stein (1990) model herding by money managers operating on incentive contracts. Lakonishok, Shleifer, Thaler, and Vishny (1991) and Chevalier and Ellison (1995) consider the possibility that money managers "window dress" their portfolios to impress investors. In two interesting recent articles, Allen and Gorton (1993) and Dow and Gorton (1994) show how money managers can churn assets to mislead their investors, and how such churning can sustain inefficient asset prices. Unlike this work, our article does not focus as much on the distortions in the behavior of arbitrageurs, as on their limited effectiveness in bringing prices to fundamental values.

The next section of the article presents a very simple model that illustrates the mechanics of arbitrage. For simplicity, our model focuses on the case where mispricing may deepen in the short run, even though there is no long run fundamental risk in the trade. We thus focus on a case that is closest to pure arbitrage, as opposed to risk arbitrage. Section II establishes the main results of the article, including our results on the effectiveness of arbitrage in extreme circumstances when prices are very far from fundamentals. Section III explores the performance-based arbitrage assumption in more detail. In section IV, we examine some empirical implications of the model. In particular, we extend the logic of the model to the more realistic case of risk arbitrage, rather than the pure arbitrage case modeled in the article. We first ask what are the characteristics of markets in which we expect risk arbitrage resources to be concentrated. We then analyze return predictability and pricing anomalies more generally. Section V concludes.

I. An Agency Model of Limited Arbitrage

The structure of the model follows Shleifer and Vishny (1990). We focus on the market for a specific asset, in which we assume there are three types of participants: noise traders, arbitrageurs, and investors in arbitrage funds who do not trade on their own. Arbitrageurs specialize in trading only in this market, whereas investors allocate funds between arbitrageurs operating in both this and many other markets. The fundamental value of the asset is V, which arbitrageurs, but not their investors, know. There are three time periods: 1, 2, and 3. At time 3, the value V becomes known to arbitrageurs and noise traders, and hence the price is equal to that value. Since the price is equal to V at $t = 3$ for sure, there is no long run fundamental risk in this trade (this is not risk arbitrage). For $t = 1, 2$, the price of the asset at time t is p_t. For concreteness, we only consider pessimistic noise traders. In each of periods 1 and 2, noise traders may experience a pessimism shock S_t, which generates for them, in the aggregate, the demand for the asset given by:

$$\mathrm{QN}(t) = [V - S_t]/p_t. \tag{1}$$

At time $t = 1$, the first period noise trader shock, S_1, is known to arbitrageurs, but the second period noise trader shock is uncertain. In particular, there is some chance that $S_2 > S_1$, i.e., that noise trader misperceptions deepen before they correct at $t = 3$. De Long et al. (1990) stressed the importance of such noise trader risk for the analysis of arbitrage.

Both arbitrageurs and their investors are fully rational. Risk-neutral arbitrageurs take positions against the mispricing generated by the noise traders. Each period, arbitrageurs have cumulative resources under management (including their borrowing capacity) given by F_t. These resources are limited, for reasons we describe below. We assume that F_1 is exogenously given, and specify the determination of F_2 below.

At time $t = 2$, the price of the asset either recovers to V, or it does not. If it recovers, arbitrageurs invest in cash. If noise traders continue to be confused, then arbitrageurs want to invest all of F_2 in the underpriced asset, since its price rises to V at $t = 3$ for sure. In this case, the arbitrageurs' demand for the asset $QA(2) = F_2/p_2$ and, since the aggregate demand for the asset must equal the unit supply, the price is given by:

$$p_2 = V - S_2 + F_2. \tag{2}$$

We assume that $F_2 < S_2$, so the arbitrage resources are not sufficient to bring the period 2 price to fundamental value, unless of course noise trader misperceptions have corrected anyway.

In period 1, arbitrageurs do not necessarily want to invest all of F_1 in the asset. They might want to keep some of the money in cash in case the asset becomes even more underpriced at $t = 2$, so they could invest more in that asset. Accordingly, denote by D_1 the amount that arbitrageurs invest in the asset at $t = 1$. In this case, $QA(1) = D_1/p_1$, and

$$p_1 = V - S_1 + D_1. \tag{3}$$

We again assume that, in the range of parameter values we are focusing on, arbitrage resources are not sufficient to bring prices all the way to fundamental values, i.e., $F_1 < S_1$.

To complete the description of the model, we need to specify the organization of the arbitrage industry and the relationship between arbitrageurs and their investors, which determines F_2. Recall that we are focusing on a particular narrow market segment in which a given set of arbitrageurs specialize. A "segment" here should be interpreted as a particular arbitrage strategy. We assume that there are many such segments and that within each segment there are many arbitrageurs, so that no arbitrageur can affect asset prices in a segment. For simplicity, we can think of T investors each with one dollar available for investment with arbitrageurs. We are concerned with the aggregate amount $F_2 \ll T$ that is invested with the arbitrageurs in a particular segment.

Arbitrageurs compete in the price they charge for their services. For simplicity, we assume constant marginal cost per dollar invested, such that all arbitrageurs in all segments have the same marginal cost. We also assume that each arbitrageur has at least one competitor who is viewed as a perfect substitute, so that Bertrand competition drives price to marginal cost. Each of the T risk-neutral investors allocates his \$1 investment to maximize expected consumer surplus, i.e., the difference between the expected return on his dollar and the price charged by the arbitrageur. Investors are Bayesians, who have prior beliefs about the expected return of each arbitrageur. Since prices are equal, an investor gives his dollar to the arbitrageur with the highest expected return according to his beliefs. Different investors hold different beliefs about various arbitrageurs' abilities, so one arbitrageur does not end up with all the funds. The market share of each arbitrageur is just the total fraction of investors who believe that he has the highest expected return. The total share of money allocated to a given segment is just the sum of these market shares across all arbitrageurs in the segment. Importantly, we assume that arbitrageurs across many segments have, on average, earned high enough returns to convince investors to invest with them rather than to index.[1]

The key remaining question is how investors update their beliefs about the future expected returns of an arbitrageur. We assume that investors have no information about the structure of the model-determining asset prices in any segment. In particular, they do not know the trading strategy employed by any arbitrageur. This assumption is meant to capture the idea that arbitrage strategies are difficult to understand, and a lot of specialized knowledge is needed for investors to evaluate them. In part, this is because arbitrageurs do not share all their knowledge with investors, and cultivate secrecy to protect their knowledge from imitation. Even if the investors were told more about what arbitrageurs were doing, they would have a difficult time deciding whether what they heard was true. Implicitly, we are assuming that the underlying structural model is sufficiently nonstationary and high dimensional that investors are unable to infer the underlying structure of the model from past returns data. As a result, they only use simple updating rules based on past performance. In particular, investors are assumed to form posterior beliefs about future returns of the arbitrageur based only on their prior and any observations of his arbitrage returns.

Under these informational assumptions, individual arbitrageurs who experience relatively poor returns in a given period lose market share to those with better returns. Moreover, since all arbitrageurs in a given segment are taking the same positions, they all attract or lose investors simultaneously, depending on the performance of their common arbitrage strategy. Specifically, investors' aggregate supply of funds to the arbitrageurs in a particular segment at time 2 is an increasing function of arbitrageurs' gross return between time 1 and time 2 (call this performance-based-arbitrage or PBA). Denoting this function by G, and recognizing that the return on the asset is given by p_2/p_1, the arbitrageurs' supply of funds at $t = 2$ is given by:

$$F_2 = F_1 * G\{(D_1/F_1) * (p_2/p_1) + (F_1 - D_1)/F_1\},$$

$$\text{with} \quad G(1) = 1, \quad G' \geq 1, \quad \text{and} \quad G'' \leq 0. \qquad (4)$$

If arbitrageurs do as well as some benchmark given by performance of arbitrageurs in other markets, which for simplicity we assume to be zero return, they neither gain nor lose funds under management. However, they gain (lose) funds if they outperform (under perform) that benchmark. Because of the

[1] See Lakonishok, Shleifer, and Vishny (1992) for a description of the agency problems in the money management industry.

extremely poor quality of investors' information, past performance of arbitrageurs completely determines the resources they get to manage, regardless of the actual opportunities available in their market.

The responsiveness of funds under management to past performance (as measured by G') is the solution to a signal extraction problem in which investors are trying to ascribe an arbitrageur's poor performance to one of three causes: 1) a random error term, 2) a deepening of noise trader sentiment (bad luck), and 3) inferior ability. High cross-sectional variation in ability across arbitrageurs will tend to increase the responsiveness of invested funds to past performance. On the other hand, if the variance of the noise trader sentiment term is high relative to the variation in (unobserved) ability, this will tend to decrease the responsiveness to past performance. In the limit, if ability is known or does not vary across arbitrageurs, poor performance could be ascribed only to a deepening of the noise trader shock (or a pure noise term), which would only increase the investor's estimate of the arbitrageur's future return. The seemingly perverse behavior of taking money away from an arbitrageur after noise trader sentiment deepens, i.e., precisely when his expected return is greatest, is a rational response to the problem of trying to infer the arbitrageur's (unobserved) ability and future opportunities jointly from past returns.

Since our results do not rely on the concavity of the G function, we focus on a linear G, given by

$$G(x) = ax + 1 - a, \qquad \text{with} \quad a \geq 1, \tag{5}$$

where x is arbitrageur's gross return. In this case, equation (4) becomes:

$$F_2 = a\ \{D_1 * (p_2/p_1) + (F_1 - D_1)\} + (1 - a)F_1 = F_1 - aD_1(1 - p_2/p_1). \tag{6}$$

With this functional form, if $p_2 = p_1$, i.e. the arbitrageur earns a zero net return, he neither gains nor loses funds under management. If $p_2 > p_1$, he gains funds and if $p_2 < p_1$, he loses funds. Note also that the higher is a, the more sensitive are the resources under management to past performance. The case of $a = 1$ corresponds to the arbitrageur not getting any more money when he loses some, whereas if $a > 1$, funds are actually withdrawn in response to poor performance.

One could in principle imagine more complicated incentive contracts that would allow arbitrageurs to signal their opportunities or abilities and attract funds based not just on past performance. For example, arbitrageurs who feel that they have superior investment opportunities might try to offer investors contracts that pay arbitrageurs a fixed price below marginal cost and a share of the upside. That is, if, at a particular point of time, arbitrageurs believe that they can earn extremely high returns with a high probability (as happens artificially at $t = 2$ in our model), they can try to attract investors by partially insuring them against further losses. We do not consider such "separating" contracts in our model, since they are unlikely to emerge in equilibrium under plausible circumstances. First, with limited liability or risk aversion, arbitrageurs might be unwilling or unable after mispricing worsens to completely retain (or increase) funds under management by insuring the investor against losses, or pricing below marginal cost. Second, these contracts are less attractive when the risk-averse arbitrageur himself is highly uncertain about his own ability to produce a superior return. We could model this more realistically by adding some noise into the third period return. In sum, under plausible conditions, the use of incentive contracts does not eliminate the effect of past performance on the market shares of arbitrageurs.[2] Empirically, most money managers in the pension and mutual fund industries work for fees proportional to assets under management and rarely get a percentage of the upside.[3] As documented by Ippolito (1992) and Warther (1995), for example, mutual fund managers lose funds under management when they perform poorly. Interestingly, Warther (1995) also shows that fund flows in and out of mutual funds affect contemporaneous returns of securities these funds hold, consistent with the results established below.

PBA is critical to our model. In conventional arbitrage, capital is allocated to arbitrageurs based on expected returns from their trades. Under PBA, in contrast, capital is allocated based on past returns, which, in the model, are low precisely when expected returns are high. At that time, arbitrageurs face fund withdrawals, and are not very effective in betting against the mispricing. Breaking the link between greater mispricing and higher expected returns perceived by those allocating capital drives our main results.

To complete the model, we need to set up an arbitrageur's optimization problem. For simplicity, we assume that the arbitrageur maximizes expected time 3 profits. Since arbitrageurs are price-takers in the market for investment services and marginal cost is constant, maximizing expected time 3 profit is equivalent to maximizing expected time 3 funds under management. For concreteness, we examine a specific form of uncertainty about S_2. We assume that, with probability q, $S_2 = S > S_1$, i.e. noise trader misperceptions deepen. With a complementary probability $1 - q$, noise traders recognize the true value of the asset at $t = 2$, so $S_2 = 0$ and $p_2 = V$.

When $S_2 = 0$, arbitrageurs liquidate their position at a gain at $t = 2$, and hold cash until $t = 3$. In this case, $W = a(D_1 * V/p_1 + F_1 - D_1) + (1 - a)F_1$.

[2] Our research assistant, Matthew Ellman of Harvard University, has solved a model in which allowing arbitrageurs to offer high-powered incentive contracts does not permit the arbitrageurs with better investment opportunities to separate themselves. The result is driven by two factors: first, limited liability precludes contracts from discouraging imitators through large penalties for poor performance, which are more likely to be levied against imitators, and, second, better arbitrageurs have more valuable alternative uses of their time, making it difficult to discourage the imitators by paying only for success since, at the contract necessary to meet the individual rationality constraint of the better arbitrageurs, the imitators still earn enough by sheer luck to cover their lower opportunity costs.

[3] Hedge fund managers typically do get a large incentive component in their compensation, but we are not aware of increases in that component, and cuts in fees, to avert withdrawal of funds.

When $S_2 = S$, in contrast, arbitrageurs third period funds are given by $W = (V/p_2) * [a\{D_1 * p_2/p_1 + F_1 - D_1\} + (1 - a)F_1]$. Arbitrageurs then maximize:

$$EW = (1 - q)\left\{a\left(\frac{D_1 * V}{p_1} + F_1 - D_1\right) + (1 - a)F_1\right\} + q\left(\frac{V}{p_2}\right) * \left\{a\left(\frac{D_1 * p_2}{p_1} + F_1 - D_1\right) + (1 - a)F_1\right\} \tag{7}$$

II. Performance-Based Arbitrage and Market Efficiency

Before analyzing the pattern of prices in our model, we specify what the benchmarks are. The first benchmark is efficient markets, in which arbitrageurs have access to all the capital they want. In this case, since noise trader shocks are immediately counteracted by arbitrageurs, $p_1 = p_2 = V$. An alternative benchmark is one in which arbitrageurs resources are limited, but PBA is inoperative, i.e., arbitrageurs can always raise F_1. Even if they lose money, they can replenish their capital up to F_1. In this case, $p_1 = V - S_1 + F_1$ and $p_2 = V - S + F_1$. Prices fall one for one with noise trader shocks in each period. This case corresponds most closely to the earlier models of limited arbitrage. There is one final interesting benchmark in this model, namely the case of $a = 1$. This is the case in which arbitrageurs cannot replenish the funds they have lost, but do not suffer withdrawals beyond what they have lost. We will return to this special case below.

The first order condition to the arbitrageur's optimization problem is given by:

$$(1 - q)\left(\frac{V}{p_1} - 1\right) + q\left(\frac{p_2}{p_1} - 1\right)\frac{V}{p_2} \geq 0 \tag{8}$$

with strict inequality holding if and only if $D_1 = F_1$, and equality holding if $D_1 < F_1$. The first term of equation (8) is an incremental benefit to arbitrageurs from an extra dollar of investment if the market recovers at $t = 2$. The second term is the incremental loss if the price falls at $t = 2$ before recovering at $t = 3$, and so they have foregone the option of being able to invest more in that case. Condition (8) holds with a strict equality if the risk of price deterioration is high enough, and this deterioration is severe enough, that arbitrageurs choose to hold back some funds for the option to invest more at time 2. On the other hand, equation (8) holds with a strict inequality if q is low, if p_1 is low relative to V (S_1 is large), if p_2 is not too low relative to p_1 (S not too large relative to S_1). That is to say, the initial displacement must be very large and prices should be expected to recover with a high probability rather than fall further. If they do fall, it cannot be by too much. Under these circumstances, arbitrageurs choose to be fully invested at $t = 1$ rather than hold spare reserves for $t = 2$. We describe the case in which mispricing is so severe at $t = 1$ that arbitrageurs choose to be fully invested as "extreme circumstances," and discuss it at some length.

This discussion can be summarized more formally in:

PROPOSITION 1: *For a given* V, S_1, S, F_1, *and* a, *there is a* q^* *such that, for* $q > q^*$, $D_1 < F_1$, *and for* $q < q^*$, $D_1 = F_1$.

If equation (8) holds with equality, the equilibrium is given by equations (2), (3), (6), and (8). If equation (8) holds with inequality, then equilibrium is given by $D_1 = F_1, p_1 = V - S_1 + F_1$, as well as equations (2) and (6). To illustrate the fact that both types of equilibria are quite plausible, consider a numerical example. Let $V = 1$, $F_1 = 0.2$, $a = 1.2$, $S_1 = 0.3$, $S_2 = 0.4$. For this example, $q^* = 0.35$. If $q < 0.35$, then arbitrageurs are fully invested and $D_1 = F_1 = 0.2$, so that the first period price is 0.9. In this case, regardless of the exact value of q, we have $F_2 = 0.1636$ and $p_2 = 0.7636$ if noise trader sentiment deepens, and $F_2 = 0.227$ and $p_2 = V = 1$ if noise trader sentiment recovers. On the other hand, if $q > 0.35$, then arbitrageurs hold back some of the funds at time 1, with the result that p_1 is lower than it would be with full investment. For example, if $q = 0.5$, then $D_1 = 0.1743$ and $p_1 = 0.8743$ (arbitrage is less aggressive at $t = 1$). If noise trader shock deepens, then $F_2 = 0.1766$, and $p_2 = 0.7766$ (arbitrageurs have preserved more funds to invest at $t = 2$), whereas if noise trader sentiment recovers then $F_2 = 0.23$ and price returns to $V = 1$. This example illustrates that both the corner solution and the interior equilibrium are quite plausible in our model. In fact, both occur for most parameters we have tried.

In this simple model, we can show that the larger are the shocks, the further are the prices from fundamental values.[4]

PROPOSITION 2: *At the corner solution* ($D_1 = F_1$), $dp_1/dS_1 < 0$, $dp_2/dS < 0$, *and* $dp_1/dS = 0$. *At the interior solution*, $dp_1/dS_1 < 0$, $dp_2/dS < 0$, *and* $dp_1/dS < 0$.

This proposition captures the simple intuition, common to all noise trader models, that arbitrageurs ability to bear against mispricing is limited, and larger noise trader shocks lead to less efficient pricing. Moreover, at the interior solution, arbitrageurs spread out the effect of a deeper period 2 shock by holding more cash at $t = 1$ and thus allowing prices to fall more at $t = 1$. As a result, they have more funds at $t = 2$ to counter mispricing at that time.

A more interesting question is how prices behave as a function of the parameter a. In particular, we would want to know whether the market becomes less efficient when PBA intensifies (a rises). Unfortunately, we do not believe that general conclusions can be drawn about how ex ante market efficiency (say, as measured by volatility) varies with a. The behavior of time 1 and time 2 prices with respect to a is very sensitive to the distribution of noise trader shocks.

In our current model, prices return to fundamentals at time 3 irrespective of the behavior of arbitrageurs. Also, the noise at time 2 either disappears or gets worse; it does not adjust part of the way toward fundamentals. Under these

[4] The proof of this proposition is straightforward, but requires some tedious calculations, which are omitted.

circumstances, we can show that a higher a makes the market less efficient. As a increases, the equilibrium exhibits the same or lower p_1 (if arbitrageurs hold back at time 1), and a strictly lower p_2 when the noise trader shock intensifies. In particular, arbitrage under PBA ($a > 0$) gives less efficient prices than limited arbitrage without PBA ($a = 0$).

On the other hand, if we modify the model to allow prices to adjust more slowly toward fundamentals, a higher a could actually make prices adjust more quickly by giving arbitrageurs more funds after a partial reversal of the noise trader shock. A partial adjustment toward fundamentals would be self-reinforcing through increased funds allocated to arbitrageurs along the way. Depending on the distribution of shocks over time, this could be the dominant effect. In general, we cannot draw any robust conclusions about ex ante market efficiency and the intensity of PBA.

However, we can say more about the effectiveness of arbitrage under extreme circumstances. In particular, we can analyze whether arbitrageurs become more aggressive when mispricing worsens. There are two ways to measure this. One is to ask whether arbitrageurs invest more total dollars in the asset at $t = 2$ than at $t = 1$, i.e., is $D_1 < F_2$? The second is whether arbitrageurs actually hold proportionally more of the asset at $t = 2$, i.e., is $D_1/p_1 < F_2/p_2$? In principle, it is possible that because $p_2 < p_1$, arbitrageurs hold more of the asset at $t = 2$ even though they spend less on it. Perhaps the clearest evidence of less aggressive arbitrage at $t = 2$ would be to show that arbitrageurs actually hold fewer shares at $t = 2$, and are liquidating their holdings, even though prices have fallen from $t = 1$. In the rest of this section, we focus on these liquidation problems.

We focus on a sufficient condition for liquidation at $t = 2$ when the noise trader shock deepens, namely, that arbitrageurs are fully invested at $t = 1$. Specifically, we have:

PROPOSITION 3: *If arbitrageurs are fully invested at $t = 1$, and noise trader misperceptions deepen at $t = 2$, then, for $a > 1$, $F_2 < D_1$ and $F_2/p_2 < D_1/p_1$.*

Proposition 3 describes the extreme circumstances in our model, in which fully invested arbitrageurs experience an adverse price shock, face equity withdrawals, and therefore liquidate their holdings of the extremely underpriced asset. Arbitrageurs bail out of the market when opportunities are the best.

Before analyzing this case in more detail, we note that full investment at $t = 1$ is a sufficient, but not a necessary condition for liquidation at $t = 2$. In general, for q's in the neighborhood above q^*, where $F_1 - D_1$ is positive but small, investors would still liquidate some of their holdings when $a > 1$. The reason is that their cash holdings are not high enough to maintain their holdings of the asset despite equity withdrawals. The cash holdings ameliorate these withdrawals, but do not eliminate them. For higher q's, however, D_1 is high enough that $F_2/p_2 > D_1/p_1$.

We can illustrate this with our numerical example from Section II, with $V = 1, S_1 = 0.3, S_2 = 0.4, F_1 = 0.2, a = 1.2$. Recall that in this example, we had

$q^* = 0.35$. One can show for this example that asset liquidations occur for $q < 0.39$, i.e., when arbitrageurs are fully invested as well as in a small region where they are not. For $q > 0.39$, arbitrageurs increase their holdings of the asset at $t = 2$.

For concreteness, it is easier to focus on the case of Proposition 3, when arbitrageurs are fully invested. In this case, we have that

$$p_2 = [V - S - aF_1 + F_1]/[1 - aF_1/p_1], \tag{9}$$

as long as $aF_1 < p_1$. The condition that $aF_1 < p_1$ is a simple stability condition in this model, which basically says that arbitrageurs do not lose so much money that in equilibrium they bail out of the market completely. If $aF_1 > p_1$, then at $t = 2$ the only equilibrium price is $p_2 = V - S$, and arbitrageurs bail out of the market completely. In the stable equilibrium, arbitrageurs lose funds under management as prices fall, and hence liquidate some holdings, but they still stay in the market.

For this equilibrium, simple differentiation yields the following result:

PROPOSITION 4: *At the fully invested equilibrium,* $dp_2/dS < -1$ *and* $d^2p_2/dadS < 0$.

This proposition shows that when arbitrageurs are fully invested at time 1, prices fall more than one for one with the noise trader shock at time 2. Precisely when prices are furthest from fundamental values, arbitrageurs take the smallest position. Moreover, as PBA intensifies, i.e., as a rises, the price decline per unit increase in S gets greater. If we think of dp_2/dS as a measure of the resiliency of the market (equal to zero for an efficient market and to -1 when $a = 0$ and there is no PBA), then Proposition 4 says that a market driven by PBA loses its resiliency in extreme circumstances. The analysis thus shows that the arbitrage process can be quite ineffective in bringing prices back to fundamental values in extreme circumstances.

This result contrasts with the more standard models, in which arbitrageurs are most aggressive when prices are furthest away from fundamentals. This point relates to Friedman's (1953) famous observation that "to say that arbitrage is destabilizing is equivalent to saying that arbitrageurs lose money on average," which is implausible. Our model is consistent with Friedman in that, on average, arbitrageurs make money and move prices toward fundamentals. However, the fact that they make money on average does not mean that they make money always. Our model shows that the times when they lose money are precisely the times when prices are far away from fundamentals, and in those times the trading by arbitrageurs has the weakest stabilizing effect.

These results are closely related to the recent studies of market liquidity (Shleifer and Vishny (1992), Stein (1995)). As in these studies, an asset here is liquidated involuntarily at a time when the best potential buyers—other arbitrageurs of this asset—have limited funds and external capital is not easily forthcoming. As a result of such fire sales, the price falls even further below fundamental value (holding the noise trader shock constant). The implication of limited resiliency for arbitrage is that arbitrage does not bring prices close to fundamental values in extreme circumstances.

The problem here may be even more severe than in operating firms. In such firms, the withdrawal/liquidation of assets is limited to the amount of debt that the firm has. In the case of arbitrage funds, unless they have a specific prohibition against withdrawals, even the equity capital can cash out because the assets themselves are liquid, as opposed to the hard assets of an operating firm. This difference in governance structures makes arbitrage funds much more susceptible to costly liquidations. In addition, investors probably understand the structure of industry downturns in operating companies better than they understand why arbitrageurs have lost their money. From this perspective as well, funds are at a greater risk of forced liquidation.

This analysis has one more interesting implication. The sensitivity to past returns of funds under management must be higher for young, unseasoned arbitrage (hedge) funds than for older, more established funds, with a long reputation for performance. As a result, the established funds will be able to earn higher returns in the long run, since they have more funds available when prices have gotten way out of line, which is when the returns to arbitrage are the greatest. In contrast, new arbitrageurs lose their funds precisely when the potential returns are the highest, and hence their average returns are lower than those of the older funds.

III. Discussion of Performance-Based Arbitrage

In our model, performance-based arbitrage, by delinking the expected return on the asset and arbitrageurs' demand for it at $t = 2$, generates the results that arbitrage is very limited. Although it is difficult to deny that PBA plays some role in the world, the question remains whether its consequences are as significant as our model suggests.

For example, one might argue that, even if funds under management decline in response to poor performance, they decline with a lag. For moderate price moves, arbitrageurs may be able to hold out and not liquidate until the price recovers. Moreover, if arbitrageurs are at least somewhat diversified, not all of their holdings lose money at the same time, suggesting again that they might be able to avoid forced liquidations.

Despite these objections, we continue to believe that, especially in extreme circumstances, PBA has significant consequences for prices. In many arbitrage funds, investors have the option to withdraw at least some of their funds at will, and are likely to do so quite rapidly if performance is poor. To some extent, this problem is mitigated by contractual restrictions on withdrawals, which are either temporary (as in the case of hedge funds that do not allow investors to take the money out for one to three years) or permanent (as in the case of closed end funds). However, these restrictions expose investors to being stuck with a bad fund manager for a long time, which explains why they are not common.[5] Moreover, creditors usually demand immediate repayment when the value of the collateral falls below (or even close to) the debt level, especially if they can get their money back before equity investors are able to withdraw their capital. Fund withdrawal by creditors is likely to be as or even more important as that by equity investors in precipitating liquidations (e.g., Orange County, December 1994). Last but not least, there may be an agency problem inside an arbitrage organization. If the boss of the organization is unsure of the ability of the subordinate taking a position, and the position loses money, the boss may force a liquidation of the position before the uncertainty works itself out. All these forces point to the likelihood that liquidations become important in extreme circumstances.

Our model shows how arbitrageurs might be forced to liquidate their positions when prices move against them. One effect that our model does not capture is that risk-averse arbitrageurs might choose to liquidate in this situation even when they don't have to, for fear that a possible further adverse price move will cause a really dramatic outflow of funds later on. Such risk aversion by arbitrageurs, which is not modeled here, would make them likely to liquidate rather than double up when prices are far away from fundamentals, making the problem we are identifying even worse. In this way, the fear of future withdrawals might have a similar effect to withdrawals themselves. We therefore expect that, even when arbitrageurs are not fully invested in a particular arbitrage strategy, significant losses in that strategy will induce voluntary liquidation behavior in extreme circumstances that looks very much like the involuntary liquidation behavior of the model.

The likelihood that risk-averse arbitrageurs voluntarily liquidate their positions in extreme circumstances is even larger if arbitrageurs are Bayesians with an imprecise posterior about the true distribution of returns on the arbitrage strategy. In that case, a sequence of poor returns may cause an arbitrageur to update his posterior and abandon his original strategy. The precision of the arbitrageur's posterior depends on the amount of past data available to estimate the return on the arbitrage strategy and on how much extra weight (if any) is placed on the more recent data. If arbitrageurs (correctly or not) believe that the world is nonstationary, they will use a shorter time series of data. This will cause their beliefs about the profitability of their strategies to be less precise (Heaton (1994)), and to change more in response to the most recent returns. This would further limit the effectiveness of arbitrage in extreme circumstances.

Finally, PBA supposes that all arbitrageurs have the same sensitivity of funds under management to performance, and that all invest in the mispriced

[5] According to the New York Stock Exchange (NYSE) Fact Book for 1993, the total dollar value of U.S. equities held by closed-end funds was only \$20.1 billion compared to \$617 billion for (open-end) mutual funds, \$1,038 billion for private pension funds (who typically have an open-end arrangement with their outside managers), and \$6,006 billion in total U.S. equities.

asset from the beginning. In fact, arbitrageurs differ. Some may have access to resources independent of past performance, and as a result might be able to invest more when prices diverge further from fundamentals. The introduction of a substantial number of such arbitrageurs can undo the effects of performance-based liquidations. If the new arbitrageurs reverse the price decline, the already invested arbitrageurs make money and hence no longer need to liquidate their holdings. However, after a very large noise trader shock that we have in the model, most arbitrageurs operating in a market are likely to find themselves fully committed. Even if some of them have held back initially, at some point most of them entered and even accumulated substantial debts to bet against the mispricing. As the mispricing gets deeper, withdrawals, as well as feared future withdrawals, cause them to liquidate. Admittedly, the total amount of capital available for arbitrage is huge, and perhaps outsiders can come in when insiders liquidate. But in practice, arbitrage markets are specialized, and arbitrageurs typically lack the experience and reputations to engage in arbitrage across multiple markets with other people's money. For this reason, outside capital does not come in to stabilize a market. In extreme circumstances, then, PBA is likely to be important and little fresh capital will be available to stabilize the market.

IV. Empirical Implications

The model presented in this article deals with the case of pure arbitrage, in which arbitrageurs do not need to bear any long run fundamental risk. While even such arbitrage must deal with problems of possible interim liquidations, in most real world situations arbitrageurs also face some long run fundamental risk. In other words, their positions pay off only on average, and not with probability one. Most data that financial economists deal with, such as stock market data, come from markets in which informed investors at best make advantageous bets. In this section, we describe some possible implications of the specialized arbitrage approach for financial markets in which arbitrageurs bear some fundamental risk, including both systematic and idiosyncratic risk. In particular, we show that this approach delivers different implications than those of noise trader models with many well-diversified arbitrageurs, such as DeLong et al (1990).

A. Which Markets Attract Arbitrage Resources?

Casual empiricism suggests that a great deal of professional arbitrage activity, such as that of hedge funds, is concentrated in a few markets, such as the bond market and the foreign exchange market. These also tend to be the markets where extreme leverage, short selling, and performance-based fees are common. In contrast, there is much less evidence of such activity in the stock market, either in the United States or abroad.[6] Why is that so? Which markets attract arbitrage?

Part of the answer is the ability of arbitrageurs to ascertain value with some confidence and to be able to realize it quickly. In the bond market, calculations of relative values of different fixed income instruments are doable, since future cash flows of securities are (almost) certain. As a consequence, there is almost no fundamental risk in arbitrage. In foreign exchange markets, calculations of relative values are more difficult, and arbitrage becomes riskier. However, arbitrageurs put on their largest trades, and appear to make the most money, when central banks attempt to maintain nonmarket exchange rates, so it is possible to tell that prices are not equal to fundamental values and to profit quickly. In stock markets, in contrast, both the absolute and the relative values of different securities are much harder to calculate. As a consequence, arbitrage opportunities are harder to identify in stock markets than in bond and foreign exchange markets.

The discussion in this article suggests a further reason why some markets are more attractive for arbitrage than others. Unlike the well-diversified arbitrageurs of the conventional models, the specialized arbitrageurs of our model might avoid extremely volatile markets if they are risk averse.

At first this claim seems counterintuitive, since high volatility may be associated with more frequent extreme mispricing, and hence more attractive opportunities for arbitrage. Assume that all volatility is due to noise trader sentiment and that the average out-performance of the arbitrageur relative to the benchmark, typically called alpha, is roughly proportional to the standard deviation of the noise trader demand shock. This means that if the arbitrageur switches to a market with twice the noise trader volatility, he also can expect twice the alpha per $1 investment. In such a market, by cutting his investment in half, the arbitrageur gets the same expected alpha and the same volatility as in the first market. He is indifferent to trading in these two markets because alpha per unit of risk is the same and he can always adjust his position to achieve the desired level of risk. This assumes that outside borrowing by the arbitrageur is limited not by the total dollar value of the investment, but by the dollar volatility of investment, which also seems plausible. In this simplified environment, the volatility of the market does not matter for the attractiveness of entry by the marginal arbitrageur.

High volatility *does*, however, make arbitrage less attractive if expected alpha does not increase in proportion to volatility. This would be true in particular when fundamental risk is a substantial part of volatility. For example, increasing one's equity position in an industry that is perceived to be underpriced carries substantial fundamental risk, and hence reduces the attractiveness of the trade. Another important factor determining the attractive-

[6] Some of these activities, such as short-selling and use of leverage, are limited by government regulations or by fund charters. Many institutions such as mutual funds are also restricted in the degree to which their positions can be concentrated in a small number of securities and in their ability to keep their positions confidential.

ness of any arbitrage concerns the horizon over which mispricing is eliminated. While greater volatility of noise trader sentiment may increase long-run returns to arbitrage, over short horizons the ratio of expected alpha to volatility may be low. Once again, this may be true for securities like equities where the resolution of uncertainty is slow and where noise trader sentiment can push prices a long way away from fundamentals before disconfirming evidence becomes available. In this case, the long run ratio of expected alpha to volatility may be high, but the ratio over the horizon of a year may be low. Markets in which fundamental uncertainty is high and slowly resolved are likely to have a high long-run, but a low short-run, ratio of expected alpha to volatility. For arbitrageurs who care about interim consumption and whose reputations are permanently affected by their performance over the next year or two, the ratio of reward to risk over shorter horizons may be more relevant. All else equal, high volatility will deter arbitrage activity.

To specialized arbitrageurs, both systematic and idiosyncratic volatility matters. In fact, idiosyncratic volatility probably matters more, since it cannot be hedged and arbitrageurs are not diversified. Ours is not the first article to emphasize that idiosyncratic risk matters in a world of information costs and specialization.[7] Merton (1987) suggests that idiosyncratic risk raises expected returns when security markets are segmented and investors must incur a fixed cost to become informed and participate in each market. Our view of risky arbitrage activity is easy to distinguish empirically from Merton's view of idiosyncratic risk in segmented markets. In Merton's model, there are no noise traders. As a result, stocks with higher idiosyncratic risk are rationally priced to earn a higher expected return. In our model, in contrast, stocks are not rationally priced, and idiosyncratic risk deters arbitrage. In particular, some stocks with high idiosyncratic variance may be overpriced, and this overpricing is not eliminated by arbitrage because shorting them is risky. These volatile overpriced stocks earn a lower expected return, unlike in Merton's model. A good example is so-called glamour stocks, or stocks of firms with higher market prices relative to various measures of fundamentals, such as earnings or book value of assets (see, for example, Lakonishok, Shleifer, and Vishny (1994)). Since these stocks have a higher than average variance of returns, a rational pricing model with segmented markets would predict higher expected returns for these stocks. In contrast, if we take the view that these stocks are overpriced, then their expected returns are lower despite the higher variance. The evidence supports the latter interpretation.

B. Anomalies

Recent research in finance has identified a number of so-called anomalies, in which particular investment strategies have historically earned higher returns than those justified by their systematic risk. One such anomaly, already

[7] The importance of idiosyncratic risk in our framework is a consequence of the assumed specialization, and not of the agency problem per se. The agency problem itself is also a natural consequence of the returns to specialization.

mentioned, is that value stocks have historically earned higher returns than glamour stocks, but there are many others. Our analysis offers a different approach to understanding these anomalies than does the standard efficient markets theory.

The efficient markets approach to these anomalies is to argue that higher returns must be compensation for higher systematic risk, and therefore the model of asset pricing that made the evidence look anomalous must have been misspecified. It must be possible to explain the anomalies away by finding a covariance between the returns on the anomalous portfolio and some fundamental factor from the intertemporal capital asset pricing model or arbitrage pricing theory.

The efficient markets approach is based on the assumption that most investors, like the economists, see the available arbitrage opportunities and take them. Excess returns are eliminated by the action of a large number of such investors, each with only a limited extra exposure to any one set of securities. Excess returns to particular securities persist only if they are negatively correlated with state variables such as the aggregate marginal utility of consumption or wealth.

As we argue in this article, the theoretical underpinnings of the efficient markets approach to arbitrage are based on a highly implausible assumption of many diversified arbitrageurs. In reality, arbitrage resources are heavily concentrated in the hands of a few investors that are highly specialized in trading a few assets, and are far from diversified. As a result, these investors care about total risk, and not just systematic risk. Since the equilibrium excess returns are determined by the trading strategies of these investors, looking for systematic risk as the only potential determinant of pricing is inappropriate. Idiosyncratic risk as well deters arbitrageurs, whether it is fundamental or noise trader idiosyncratic risk.

Our article suggests a different approach to understanding anomalies. The first step is to understand the source of noise trading that might generate the mispricing in the first place. Specifically, it is essential to examine the demand of the potential noise traders, whether such demand is driven by sentiment or institutional restrictions on holdings. The second step is to evaluate the costs of arbitrage in the market, especially the total volatility of arbitrage returns. For a given noise trading process, volatile securities will exhibit greater mispricing and a higher average return to arbitrage in equilibrium. (Other costs of arbitrage, such as transaction costs, are also important (Pontiff (1996)).

We can illustrate the difference between the two approaches using the value/glamour anomaly. To justify an efficient markets approach to explaining this anomaly, Fama and French (1992) argue that the capital asset pricing model is misspecified, and that high (low) book to market stocks earn a high (low) return because the former have a high loading on a different risk factor than the market. Although they don't precisely identify a macroeconomic factor to which the high book to market stocks are particularly exposed, they argue that the portfolio of high book to market stocks is itself a proxy for such a factor, which they call the distress factor.

Our approach instead would be to identify the pattern of investor sentiment responsible for this anomaly, as well as the costs of arbitrage that would keep it from being eliminated. To begin, the glamour-value evidence is consistent with some investors extrapolating past earnings growth of companies and failing to recognize that extreme earnings growth is likely to revert to the mean (Lakonishok, Shleifer, and Vishny (1994), LaPorta (1996)). With respect to risk, the conventional arbitrage of the glamour-value anomaly, i.e., simply taking a long position in a diversified portfolio of value (high book-to-market) stocks, has been roughly a 60–40 proposition over a one year horizon. That is, the odds of outperforming the S&P 500 index over one year have been only 60 percent, although over 5 years the superior performance has been much more likely.[8] Over a short horizon, then, arbitrage returns on the value portfolio are volatile. Even though this risk may be idiosyncratic, it cannot be hedged by arbitrageurs specializing in this segment of the market. Because of the high volatility of the hedge strategy, and the relatively long horizon it relies on to secure positive returns with a high probability, it is likely to be shunned by arbitrageurs, particularly those with a short track record.

Our approach further implies that, in extreme situations, arbitrageurs trying to eliminate the glamour/value mispricing might lose enough money that they have to liquidate their positions. In this case, arbitrageurs may become the least effective in reducing the mispricing precisely when it is the greatest. Something along these lines occurred with the stocks of commercial banks in 1990–1991. As the prices of these stocks fell sharply, many traditional value arbitrageurs invested heavily in these stocks. However, the prices kept falling, and many value arbitrageurs lost most of their funds under management. As a consequence, they had to liquidate their positions, which put further pressure on the prices of banking stocks. After this period, the returns on banking stocks have been very high, but many value funds did not last long enough to profit from this recovery.

The glamour/value anomaly is one of several that our approach might explain. The analysis actually predicts what types of market anomalies can persist over the long term. These anomalies must have a high degree of unpredictability, which makes betting against them risky for specialized arbitrageurs. However, unlike in the efficient markets model, this risk need not be correlated with any macroeconomic factors, and can be purely idiosyncratic fundamental or noise trader risk.

Finally, the specialized arbitrage approach assumes that only a relatively small number of specialists understand the return anomaly well enough to exploit it. This may be questionable in the case of anomalies like the value-glamour anomaly or the small firm anomaly about which there is now much published work. As more investors begin to understand an anomaly, the superior returns to the trading strategy may be diminished by the actions of a larger number of investors who each tilt their portfolios toward the underpriced assets. Alternatively, investors may become more knowledgeable about

[8] The exact odds depend on what sample period and what universe of stocks is used.

the strategies being used and judge arbitrageurs relative to a more accurate benchmark of their peers (e.g., other value managers or a value index), thereby diminishing some of the withdrawals when an entire peer group is performing poorly. The specialized arbitrage approach is clearly more appropriate for difficult-to-understand new arbitrage opportunities than it is for well-understood anomalies (which should presumably not be anomalies for long).

We would nonetheless argue that anomalies become understood very slowly and that investors do not take definitive action on their information until long after a phenomenon has been exposed to public scrutiny. The anomaly is more easily accepted when the pattern of returns is not very noisy and the payoff horizon is short (such as the small firm effect in January). A "noisy" anomaly like the value-glamour anomaly is accepted only slowly, even by relatively sophisticated investors.

V. Conclusion

Our article describes the workings of markets in which specialized arbitrageurs invest the capital of outside investors, and where investors use arbitrageurs' performance to ascertain their ability to invest profitably. We show that such specialized performance-based arbitrage may not be fully effective in bringing security prices to fundamental values, especially in extreme circumstances. More generally, specialized, professional arbitrageurs may avoid extremely volatile "arbitrage" positions. Although such positions offer attractive average returns, the volatility also exposes arbitrageurs to risk of losses and the need to liquidate the portfolio under pressure from the investors in the fund. The avoidance of volatility by arbitrageurs also suggests a different approach to understanding persistent excess returns in security prices. Specifically, we expect anomalies to reflect not some exposure of securities to difficult-to-measure macroeconomic risks, but rather, high idiosyncratic return volatility of arbitrage trades needed to eliminate the anomalies. In sum, this more realistic view of arbitrage can shed light on a variety of observations in securities markets that are difficult to understand in more conventional models.

REFERENCES

Allen, Franklin, 1990, The market for information and the origin of financial intermediation, *Journal of Financial Intermediation* 1, 3–30.

Allen, Franklin, and Gary Gorton, 1993, Churning bubbles, *Review of Economic Studies* 60, 813–836.

Bhattacharya, Sudipto, and Paul Pfleiderer, 1985, Delegated portfolio management, *Journal of Economic Theory* 36, 1–25.

Campbell, John, and Albert Kyle, 1993, Smart money, noise trading, and stock price behavior, *Review of Economic Studies* 60, 1–34.

Chevalier, Judith, and Glenn Ellison, 1995, Risk taking by mutual funds as a response to incentives, manuscript.

DeLong, J. Bradford, Andrei Shleifer, Lawrence Summers, and Robert Waldmann, 1990, Noise trader risk in financial markets, *Journal of Political Economy* 98, 703–738.

Dow, James, and Gary Gorton, 1994, Noise trading, delegated portfolio management, and economic welfare, NBER Working paper 4858.

Fama, Eugene, 1965, The behavior of stock market prices, *Journal of Business* 38, 34–105.

Fama, Eugene, and Kenneth French, 1992, The cross-section of expected stock returns, *Journal of Finance* 46, 427–466.

Friedman, Milton, 1953, The case for flexible exchange rates, in *Essays in Positive Economics* (University of Chicago Press, Chicago).

Grossman, Sanford, and Merton Miller, 1988, Liquidity and market structure, *Journal of Finance* 43, 617–633.

Heaton, John, 1994, Learning and the belief in low-scaled price portfolio strategies, 1940–1993, manuscript, University of Chicago.

Ippolito, Richard, 1992, Consumer reaction to measures of poor quality: evidence from the mutual fund industry, *Journal of Law and Economics* 35, 45–70.

LaPorta, Rafael, 1996, Expectations and the cross-section of stock returns, *Journal of Finance* 51, 1715–1742.

Lakonishok, Josef, Andrei Shleifer, Richard Thaler, and Robert Vishny, 1991, Window dressing by pension fund managers, *American Economic Review Papers and Proceedings* 81, 227–231.

Lakonishok, Josef, Andrei Shleifer, and Robert Vishny, 1992, The structure and performance of the money management industry, *Brookings Papers on Economic Activity: Microeconomics*, 339–391.

Lakonishok, Josef, Andrei Shleifer, and Robert Vishny, 1994, Contrarian investment, extrapolation, and risk, *Journal of Finance* 49, 1541–1578.

Merton, Robert, 1987, A simple model of capital market equilibrium with incomplete information, *Journal of Finance* 42, 483–510.

Pontiff, Jeffrey, 1996, Costly arbitrage: Evidence from closed-end funds. *Quarterly Journal of Economics* 111, 1135–1152.

Ross, Steven, 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.

Scharfstein, David, and Jeremy Stein, 1990, Herd behavior and investment, *American Economic Review* 80, 465–489.

Sharpe, William, 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425–442.

Sharpe, William, and Gordon Alexander, 1990, Investments, 4th edition, (Prentice Hall, Englewood Cliffs, N.J.).

Shleifer, Andrei, and Robert Vishny, 1990, Equilibrium short horizons of investors and firms, *American Economic Review Papers and Proceedings* 80, 148–153.

Shleifer, Andrei, and Robert Vishny, 1992, Liquidation values and debt capacity: A market equilibrium approach, *Journal of Finance* 47, 1343–1366.

Stein, Jeremy, 1995, Prices and trading volume in the housing market: A model with downpayment effects, *Quarterly Journal of Economics*, 110, 379–406.

Warther, Vincent, 1995, Aggregate mutual fund flows and security returns, *Journal of Financial Economics* 39, 209–236.