NBER WORKING PAPER SERIES

THE LIMITS OF ARBITRAGE

Andrei Shleifer
Robert W. Vishny

Working Paper No. 5167

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 1995

Nancy Zimmerman and Gabe Sunshine have helped us understand arbitrage. We also thank Yacine Ait-Sahalia, Douglas Diamond, Oliver Hart, Steve Kaplan, Raghu Rajan, Jesus Saa-Requejo, and Luigi Zingales for helpful comments. This paper is part of NBER's research program in Asset Pricing. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #5167
July 1995

THE LIMITS OF ARBITRAGE

ABSTRACT

In traditional models, arbitrage in a given security is performed by a large number of diversified investors taking small positions against its mispricing. In reality, however, arbitrage is conducted by a relatively small number of highly specialized investors who take large positions using other people's money. 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.

Andrei Shleifer
Department of Economics
Harvard University
Cambridge, MA 02138
and NBER

Robert W. Vishny
Graduate School of Business
University of Chicago
1101 East 58th Street
Chicago, IL 60637
and NBER

The Limits of Arbitrage[1]

Andrei Shleifer and Robert W. Vishny

Harvard University and University of Chicago

May 1995

I. Introduction

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). In standard models of efficient markets, the workings of arbitrage brings prices close to fundamental values. The simplest forms of arbitrage, such as cash futures arbitrage or replication of an option, are literally riskless, and are limited only by transaction costs. In practice, however, the securities arbitrageurs trade are similar rather than the same, and the arbitrage they engage in is risky. Even in models such as CAPM and APT, arbitrageurs must bear some idiosyncratic risk in their trading strategies.

One model of risky arbitrage is that of a large number of investors taking small positions against the mispricing. Fama's (1965) classic analysis of efficient markets and Ross's (1976) Arbitrage Pricing Theory are based on this model. An alternative, and in many cases more realistic, view is that arbitrage is conducted by a relatively few professional, highly specialized investors who combine their knowledge with resources of outside investors to take large positions. They operate in markets where fundamentals are

[1]Nancy Zimmerman and Gabe Sunshine have helped us understand arbitrage. We also thank Yacine Ait-Sahalia, Douglas Diamond, Oliver Hart, Steve Kaplan, Raghu Rajan, Jesus Saa-Requejo, and Luigi Zingales for helpful comments.

difficult to ascertain and correct hedging strategies are hard to implement, such as the currency and derivative markets. 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 paper, we examine such arbitrage and its effectiveness in achieving market efficiency.

The principal question we are interested in is: how effective is professional arbitrage in extreme circumstances, when prices are far away from fundamental values? Intuition suggests that the greater the deviation of prices from fundamental values in the absence of arbitrage, the greater the positions that arbitrageurs would take to counter these deviations, and hence the more effective is arbitrage. Put simply, greater mispricing attracts more arbitrage resources. Indeed, the available models of risky arbitrage, such as Grossman and Miller (1988), De Long, Shleifer, Summers and Waldmann (1990), or Campbell and Kyle (1992) all share this feature. In these models, arbitrage is not perfect because arbitrageurs face either fundamental or noise trader risk (the risk that mispricing becomes more extreme). Nonetheless, this risk does not outweigh the higher expected returns when mispricing is greater. At a substantive level, the implication is that arbitrage is stabilizing, and it is more stabilizing in extreme circumstances.

In the analysis that follows, we examine this logic in the agency context of the arbitrage activity. The separation of knowledge and resources has three crucial implications for the workings of arbitrage. First, because outside investors are ignorant about the markets that arbitrageurs invest in and cannot tell good arbitrageurs from bad, the resources they supply to the arbitrage activity are limited. Arbitrageurs can perhaps borrow funds in addition to resources they have under management, but their borrowing capacity is limited as well. Second, again because they are poorly informed, outside investors rationally use past performance of the arbitrageurs to gauge their ability and to increase and reduce funds they give them to manage. A good track record brings in more funds, and a bad track record causes a withdrawal of funds. Third, because arbitrageurs' knowledge is highly specialized, arbitrage markets are segmented. Only a relatively few experts, with a good track record, can attract outside funds to engage in arbitrage in a given market.

These simple implications of agency have important consequences for arbitrage in extreme circumstances, when prices diverge significantly from fundamental values. Such price divergences can occur, for example, because unsophisticated investors experience a sharp move in their sentiment, and as a result try to significantly change their holdings. They may become highly enthusiastic about particular securities, or they may panic. In this paper, we focus not on the origins of unpredictable changes in investor sentiment, but rather on arbitrageurs' response. We show that, in the agency context of specialized arbitrage using outside funds, arbitrage can be very ineffective in returning prices to fundamental values when investor sentiment has driven them far away.

Ours is obviously not the first study of the consequences of delegated portfolio management. Early papers in this area include Allen (1990) and Bhattacharya-Pfleiderer (1985). Scharfstein and Stein (1990) have modelled herding by money managers operating on incentive contracts. Lakonishok, Shleifer, Thaler and Vishny (1991) and Chevalier and Ellison (1995) have considered the possibility that money managers "window dress" their portfolios to impress investors. In two interesting recent papers, Allen and Gorton (1993) and Dow and Gorton (1994) have shown how money managers can churn assets to mislead their investors, and how such churning can sustain inefficient asset prices. Unlike this work, our paper 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 paper presents a very simple model that illustrates the mechanics of specialized arbitrage. This model shows why arbitrage cannot be counted on to bring security prices close to fundamental values when they have moved far away. The model also demonstrates how markets become less resilient during the times of extreme mispricing.

In section 3, we examine some implications of this model. We first ask what are the characteristics of markets in which we expect arbitrage resources to be concentrated. We then analyze the implications of the model for return predictability and pricing anomalies more generally.

Section 4 concludes.

2. An agency model of limited arbitrage.

Specification

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. 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:

(1) $QN(t) = [V - S_t] / p_t$.

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:

(2) $p_2 = V - S_2 + F_2$.

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

(3) $p_1 = V - S_1 + D_1$.

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. 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. Competition between arbitrageurs ensures that the price they charge for their services is driven down to this 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, and 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[2].

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 fact, they do not even know which arbitrageurs trade in which market segments. 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

---

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

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 non-stationary 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' 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:

(4) $F_2 = G\{D_1*(p_2/p_1) + (F_1 - D_1)\}$, with $G' > 1$ and $G'' < 0$.

If arbitrageurs do as well as some benchmark given by performance of arbitrageurs in other markets, they neither gain nor lose funds under management. However, they gain (lose) funds if they outperform (underperform) that benchmark. Because of the 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.

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 with a fixed price below marginal cost which give themselves 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. 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. Loss of money under management is the most immediate consequence when these managers perform poorly. Hedge fund managers typically do get a large incentive component in their compensation, but they are also paid high fixed fees.

PBA is critical to our model. For arbitrage to be most effective in situations of extreme mispricing, capital must be allocated to arbitrageurs based on expected returns from their arbitrage 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 are at their leverage constraint and face equity withdrawals, and so are the least 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 calculate the equilibrium D_1, p_1 and p_2, we need to set up an arbitrageur's optimization problem. Since managers are paid fees equal to the marginal cost per dollar of funds under management, their welfare is an increasing function of their expected third period funds under management (market share). These funds are given by:

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

i.e. the product of the return on period 2 investment and the size of that investment, F_2.

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, $S_2 = 0$ 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, $W = G\{D_1* V/p_1 + (F_1 - D_1)\}$. When $S_2 = S$, in contrast, arbitrageurs third period funds are given by $W = (V/p_2)*G\{D_1*p_2/p_1 + (F_1 - D_1)\}$. Arbitrageurs then maximize:

(5) $EW = (1\text{-}q)G\{D_1* V/p_1 + (F_1 - D_1)\}+ q(V/p_2)*G\{D_1*p_2/p_1 + (F_1 - D_1)\}$.

The first order condition is given by:

$$(6)\ (1 - q)G'\left(D_1\frac{V}{p_1} + F_1 - D_1\right)\left(\frac{V}{p_1} - 1\right) + qG'\left(D_1\frac{p_2}{p_1} + F_1 - D_1\right)\left(\frac{p_2}{p_1} - 1\right)\frac{V}{p_2} = 0$$

We are interested in the circumstances in which arbitrageurs fully invest at t = 1, i.e. they choose $D_1 = F_1$. This will happen if (6) is positive even at $D_1 = F_1$, i.e., if

$$(7)\ (1 - q)G'\left(F_1\frac{V}{p_1}\right)\left(\frac{V}{p_1} - 1\right) + qG'\left(F_1\frac{p_2}{p_1}\right)\left(\frac{p_2}{p_1} - 1\right)\frac{V}{p_2} > 0$$

The first term of (7), which is positive, is an incremental benefit to arbitrageurs, at $D_1 = F_1$, from an extra dollar of investment if the market recovers at t=2. The second term, which is negative, 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. It is easy to see that (7) holds 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), and if G is not too concave. 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. Moreover, the penalty for underperformance cannot be too severe relative to reward for outperformance. Under these circumstances, condition (7) holds and so arbitrageurs choose to be fully invested at t=1 rather than hold spare reserves for t=2.

Condition (7) clarifies the sense in which our model deals with the workings of arbitrage in extreme circumstances. We are looking at a situation where an extreme noise trader shock (or liquidation) has occurred, and arbitrageurs do not expect a much greater shock to follow. As a result, they choose not to maintain extra reserves, but to bet all their funds on a price recovery. The option value of waiting for prices to fall further is not very high. Nonetheless, there is some chance that prices fall even further because noise traders misperceptions deepen, and arbitrageurs are caught with poor performance. This is the scenario that our model sheds light on.

Condition (7) describes the most interesting case, but the model can also be analyzed when (7) fails, and arbitrageurs choose an interior D_1. In this case, if noise traders' misperceptions deepened at t=2, arbitrageurs would lose money under management and the price would fall at t=2, but arbitrageurs investment would still rise ($F_1 > F_2 > D_1$). Arbitrageurs in this case bear more heavily against the mispricing in the extreme circumstances, just as in the usual models.

When condition (7) holds, the solution for the second period price is given implicitly by:

(8) $p_2 = V - S + G(F_1 * p_2/p_1)$.

There is a stable equilibrium if G′ is not too high, i.e., $G'*F_1/p_1 < 1$. [Otherwise, when prices fall at t=2, funds under management fall so much as to require an even further drop in prices, making $p_2 = V-S$ the only sustainable equilibrium, in which arbitrage completely collapses at t=2. In this case, however, it is less plausible to assume that (7) holds]. In the stable equilibrium, arbitrageurs expect prices to fall if noise traders misperceptions deepen, and money under management to decline as prices fall. Moreover, the price decline is consistent with equilibrium funds that arbitrageurs in this segment end up managing.

The interesting aspect of this equilibrium is sensitivity of p_2 with respect to second period noise trader shock S. Differentiating (8) with respect to S and rearranging terms, we obtain:

(9) $$\frac{dp_2}{dS} = \frac{-1}{1 - G'\frac{F_1}{p_1}} < -1$$

Since in a stable equilibrium $G'*F_1/p_1 < 1$, p_2 falls more than one for one with second period noise trader shock S. In the extreme situation where noise trader shocks deepen starting from an already bad situation, arbitrageurs end up reducing their demand for the underpriced asset, and the price falls even further than it would without this adverse shift in arbitrageurs' demand.

In this simple example, in the unlikely event that noise trader misperceptions deepen even after they are already extreme, arbitrageurs withdraw their support of the price. They are already fully invested, but lose funds under management and so have to liquidate their holdings. Even though prices are far below fundamental values, arbitrageurs are trying to sell and prices fall more than one for one with the noise trader shock. Precisely when prices are furthest from fundamental values, arbitrageurs take the smallest position. The example thus shows that the arbitrage process can be quite ineffective in bringing prices back to fundamental values in extreme circumstances.

In this model, the derivative of the price with respect to noise trader shock measures the resiliency of the market. If markets were perfectly efficient, this measure would be zero. At t=1, by construction of the model, since arbitrage funds are limited, this measure is 1: prices fall one for one with noise trader shock. What is perhaps most interesting here is that if noise trader sentiment does not recover, prices fall more than one for one at t=2. In these extreme circumstances, the marginal trades of the arbitrageurs make the market less resilient solely because of how arbitrage resources are allocated by their investors.

These results are closely related to the recent studies of market liquidity (Shleifer and Vishny 1992, Kiyotaki and Moore 1994, Stein 1995). As in these studies, an asset here is liquidated involuntarily at a time when the best potential buyers, namely 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 model has one more interesting implication. The sensitivity of funds under management to past returns 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 funds available during the times of extreme deviations of prices from fundamental values, 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.

Discussion of Performance Based Arbitrage

In our model, performance-based-arbitrage breaks the link between the expected return on the asset and arbitrageurs' demand for it at $t = 2$, and hence generates our key result that arbitrage is ineffective in extreme circumstances, precisely when returns to arbitrage are the highest. Is PBA plausible?

In some situations, funds under management do not decline very fast when performance is only moderately poor and hence the liquidation effect we are emphasizing is likely to be small. Indeed, the arbitrage community is likely to avoid significant liquidation for small noise trader shocks, since G' is likely to be close to one in the neighborhood of F_1. Moreover, when prices have moved only moderately from fundamental values, arbitrageurs are likely not to be fully invested. That is precisely why our model is most applicable in extreme circumstances, where many of the arbitrageurs operating in a market are already fully invested, experience substantial losses and face fund withdrawals and liquidations.

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. However, 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 in extreme circumstances. 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. 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.

The arguments above explain 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 modelled 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.

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 non-stationary, 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 are the same in terms of sensitivity of funds under management to performance, and that all invest in the mispriced 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 need to 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 hold and little fresh capital will be available to stabilize the market.

Summary

In our model, we made three key assumptions about agency problems in arbitrage. First, in extreme circumstances, arbitrageurs are already fully invested: they are at their leverage constraint and have no access to fresh equity capital. Second, fresh equity capital is allocated based on past performance because investors in arbitrage funds do not know the structure of the model. Third, markets are segmented, and no new equity or debt capital flows to arbitrage activities from outside investors who are both sophisticated and not already invested. Under these assumptions, arbitrage fails to bring asset prices back to fundamental values when they have moved far away. Arbitrageurs lean against the mispricing, but their ability to do so is very limited. This result contrasts with the more standard models, in which arbitrageurs are most aggressive when prices are furthest away from fundamentals.

Another way to make this point is to relate it 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.

3. Implications.

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. 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. In foreign exchange markets, calculations of relative values are more difficult. However, arbitrageurs put on their largest trades, and make the most money, when central banks attempt to maintain non-market exchange rates, in which case it is possible to tell that prices are not equal to fundamental values and to profit quickly. In contrast, in stock markets, 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 paper suggests a different 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 to begin that all volatility is due to noise trader sentiment and that the average outperformance 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 attractiveness 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 resolved slowly 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 other things equal, then, 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. Merton (1986) has suggested that idiosyncratic risk raises expected returns when security markets are segmented. 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.

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 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. The anomalies must be possible to explain 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.

Theoretically, 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 argued in this paper, 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 funds care about total risk, and not just systematic risk. Since the equilibrium excess returns are determined by the trading strategies of these arbitrageurs, 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 paper suggests a new 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 1994)].

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, 1994) 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. They next need to find the macroeconomic factor to which the high book to market stocks are particularly exposed. Although they don't precisely identify such a factor, they argue that the portfolio of high book to market stocks must proxy for such a factor, which they choose to call the distress factor. They further find that returns to the value (or glamour) stocks seem to have a common component, which is captured by their mimicking portfolio. They find this co-movement to be evidence supportive of their fundamental risk interpretation of the anomaly.

The absence of clearly interpretable macroeconomic variables capable of explaining average returns on the Fama-French portfolios is troubling for the efficient markets interpretation of the evidence. Moreover, as De Long et al (1990) show, co-movement between large numbers of securities per se is not necessarily supportive of the efficient markets approach. In the value/glamour context, there are many years in which value stocks as a group have done significantly better or worse than glamour stocks. This can easily be explained by changes in investor sentiment about the relative future returns on these two groups of stocks. The key question is whether the value minus glamour return is negatively correlated with macroeconomic state variables such as the (aggregate) marginal utility of consumption or wealth. This would be necessary to justify the higher equilibrium returns to value stocks. Evidence in LSV (1994) indicates that, if anything, the correlation between value minus glamour returns and the marginal utility of consumption or wealth is positive. The traditional asset pricing approach is hard pressed to explain these facts.

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 arbitraged away. In LSV (1994), we argued that the glamour-value evidence is consistent with some investors extrapolating past performance of companies and failing to recognize that extreme performance is likely to revert to the mean. That paper, as well as La Porta (1993) and La Porta et al (1994), has presented evidence indicating that the pricing of glamour and value stocks is consistent with extreme but inaccurate expectations about their future performance. For example, investors in glamour stocks appear to be consistently disappointed in their earnings performance, suggesting that their initial forecasts were too optimistic. The converse is true for value stocks.

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[3]. Over a short horizon, then, arbitrage returns on the value portfolio are volatile. And 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 has 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 obviously one of many that we believe can be explained using our analysis. In fact, the analysis actually predicts what types of anomalies can persist in the market over the long term. These anomalies must have a high degree of unpredictability, which makes arbitraging them away risky for specialized arbitrageurs. At the same time, 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.

---

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

4. Conclusion.

In traditional finance models, arbitrage of a mispriced security is performed by a large number of diversified investors taking small positions in it. In reality, arbitrage is performed by relatively few highly specialized and undiversified investors using other peoples' funds. In this agency context, the implications of arbitrage for security pricing are very different from the standard results.

Our paper has shown 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 new 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. F., 1990, "The market for information and the origin of financial intermediation," Journal of Financial Intermediation 1: 3-30.

Allen, F. and G. Gorton, 1993, Churning bubbles, Review of Economic Studies 60: 813-836.

Bhattacharya, S. and P. Pfleiderer, 1985, "Delegated portfolio management," Journal of Economic Theory, 36: 1-25.

Campbell, J. and A. Kyle, 1993, Smart money, noise trading, and stock price behavior, Review of Economic Studies 60: 1-34.

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

De Long, J.B, A. Shleifer, L. Summers, and R. Waldmann, 1990, Noise trader risk in financial markets, Journal of Political Economy 98, 703-738.

Dow, J. and G. Gorton, 1994, Noise trading, delegated portfolio management, and economic welfare, NBER working paper 4858.

Fama, E., 1965, The behavior of stock market prices, Journal of Business 38: 34-105.

Fama, E. and K. French, 1992, The cross-section of expected stock returns, Journal of Finance 46, 427-466.

Fama, E. and K. French, 1993, Size and book-to-market factors in earnings and returns, manuscript.

Friedman, M., 1953, The case for flexible exchange rates, in Essays in Positive Economics, Chicago: University of Chicago Press.

Grossman, S. and M. Miller, 1988, Liquidity and market structure, Journal of Finance 43, 617-633.

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

Kiyotaki, N. and J. Moore, 1994, Credit Cycles, Mimeo, London School of Economics.

La Porta, R., 1993, Expectations and the cross-section of stock returns, manuscript.

La Porta, R., J. Lakonishok, A. Shleifer, and R. Vishny, 1995, Good news for value stocks: further evidence on market efficiency, manuscript.

Lakonishok, J., A. Shleifer, R. Thaler, and R. Vishny, 1991, Window dressing by pension fund managers, American Economic Review Papers and Proceedings 81: 227-231.

Lakonishok, J., A. Shleifer, and R. Vishny, 1992, The structure and performance of the money management industry, Brookings Papers on Economic Activity: Microeconomics, 339-391.

Lakonishok, J., A. Shleifer, and R. Vishny, 1994, Contrarian investment, extrapolation, and risk, Journal of Finance 49, 1541-1578.

Merton, R.C., 1987, A simple model of capital market equilibrium with incomplete information, Journal of Finance 42, 483-510.

Myers, S. and N. Majluf, 1984, Corporate financing and investment decisions when firms have information that investors do not have, Journal of Financial Economics 13, 187-221.

Pontiff, J., 1994, Costly arbitrage, mispricing, and closed-end fund discounts, manuscript.

Ross, S., 1976, The arbitrage theory of capital asset pricing, Journal of Economic Theory 13: 341-360.

Scharfstein, D. and J. Stein, 1990, Herd behavior and investment, American Economic Review 80: 465-489.

Sharpe, W. and G. Alexander, 1990, Investments, 4th edition, Englewood Cliffs, N.J.: Prentice Hall.

Shleifer, A. and R. Vishny, 1990, Equilibrium short horizons of investors and firms, American Economic Review Papers and Proceedings 80, 148-153.

Shleifer, A. and R. Vishny, 1992, Liquidation values and debt capacity: a market equilibrium approach, Journal of Finance 47, 1343-1366.

Stein, J., 1995, Prices and trading volume in the housing market: a model with downpayment effects, Quarterly Journal of Economics, CX: 379-406.