

# Limited Arbitrage, Segmentation, and Investor Heterogeneity: Why the Law of One Price So Often Fails

Sean Masaki Flynn\*

August 28, 2003

## ABSTRACT

There are numerous examples of assets with identical payout streams being priced differently. These violations of the law of one price result from two factors. First, investors have heterogeneous asset valuations so that if two groups of investors trade in segmented markets they are likely to set different prices because they have different expectations as to the value of the identical assets. Second, such discrepancies can only persist if arbitrage activities are limited. There appear to be two major limitations, short sales constraints and noise trader risk. Those assets facing short sales constraints have an asymmetric distribution of pricing violations because short sales constraints only bind when asset prices are too high. By contrast, assets facing noise trader risk have symmetric violation distributions because noise trader risk must be born by arbitrageurs both when prices are too low as well as too high.

---

\*Department of Economics, Vassar College, 124 Raymond Ave. #424, Poughkeepsie, NY 12604. flynn@vassar.edu I would like to thank Elias Khalil, the Behavioral Research Council, and the American Institute for Economic Research for their support. I am also indebted to George Akerlof, Steve Ross, and Ernst Fehr for their encouragement.

# I. Introduction

The orthodox Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965) rests upon the assumption that Sharpe calls “homothetic expectations.” This is the idea that rational investors, if exposed to the same information set, will come to the same estimate about a given asset’s future payout distribution. Consequently, the only differences in behavior that we will see among investors will be the result of differences in their respective risk aversions. Indeed, in the CAPM world, there is no disagreement about asset prices, as all investors, seeing the same information, price each asset in the same way—on the basis of how much marginal risk a marginal amount of an asset adds to the total riskiness of a diversified portfolio.

Three well know conclusions follow from the CAPM model. First there should be very low volume seen on stock markets. Second, all investors will choose to hold the market portfolio. Third, the rate of return on an asset should be governed by its correlation with the market portfolio.

Each of these three famous predictions of the CAPM model fails spectacularly in practice. There is extremely high volume in the asset markets. Very few investors choose to buy the market portfolio. And, at least *ex post*, asset returns are very poorly predicted by their correlation with the return on the market portfolio.

Miller (1977) presents a very different mechanism for the determination of asset prices, a model which can explain the failure of the three CAPM predictions.<sup>1</sup> However, the focus of this paper will be on the ability of the Miller (1977) model to explain why financial markets feature so many instances of the violation of the most fundamental implication of rational asset pricing theory, the law of one price. Each violation involves two assets that trade at different prices despite having identical future payout streams. As will be discussed below, such arbitrage violations are impossible in the presence of homothetic investors, but are more

likely than not in the Miller (1977) model's world of heterogeneous investors segmented into different markets.

The greater data availability of recent years has lent strong empirical support to Miller (1977). Additionally, formal models including the two-period model of Chen, Harrison, and Stein (2002) and the full intertemporal treatments of Duffie, Garleanu, and Pedersen (2002) and Gallmeyer and Hollifield (2002) will hopefully convince the profession that Miller's insight, now that it has been made mathematically precise, is a robust alternative to the CAPM orthodoxy which has aged poorly.

The remainder of this paper is organized as follows. Section II discusses Miller's model, its subsequent formalization by other authors, and the evidence in support of it. Section III demonstrates that Miller's model can explain each of the most famous violations of the law of one price, including the ARCO/Exxon price divergence, the Royal Dutch/Shell price divergence, the Palm/3Com carve-out anomaly, the asymmetric failure of put-call parity, and the existence and behavior of discounts on closed-end mutual funds. Section IV discusses the relationship between the definiteness of the arbitrage horizon and the willingness of arbitrageurs to risk capital when encountering violations of the law of one price. Section V concludes.

## **II. The Miller Model of Asset Pricing**

Miller (1977) views asset pricing as an auction in which the finite number of shares issued by a company go to those investors having the most optimistic expectations about the company's future profitability. As with most auctions, this leads to a winner's curse, with those most optimistic often overpaying for the asset.

Miller's model is predicated upon investors having heterogeneous expectations. By contrast, traditional asset pricing assumes that each rational investor, having access to the same information set, will come to hold the same beliefs about an asset's future payout stream as every other investor. This consensus implies that they will agree on the price at which that payout stream should be currently valued. Miller (1977) disputes this, and in doing so produces a simple model which can account for not only the failings of CAPM but the much more fundamental issue of why the law of one price so often fails.

Assume that there are  $N$  investors. For expositional convenience, let each investor be limited to one-share positions in the stock of a given company. That is, they can either go long one share, go short one share, or hold no shares at all. Assume that the investors have heterogeneous beliefs that are common knowledge.<sup>2</sup> Next, rank the investors from highest to lowest, first in terms of their beliefs about the expected returns on the asset, and then again in terms of the current valuations they assign to the asset. This will give you both a distribution of investor demands as well as the market demand curve for the investment.

Figure 1 shows the investors' demand distribution. Those who have extremely positive expectations demand a few shares on the right end of the distribution, while the more numerous investors having more moderate expectations demand more shares towards the middle of the distribution. If the total number of shares available is given by  $F$ , then all those shares will be bought up by the investors having the most optimistic expectations about returns. In Figure 1, the vertically lined right tail ends up with all the  $F$  shares as these investors will outbid other investors for the ownership rights.

The price determination of the asset can be seen in Figure 2, where investors are lined up, left to right, from highest valuation to lowest valuation. This traces out a demand curve. Inserting a vertical supply curve at  $F$  units gives the market price of the stock at the intersection of the vertical supply curve with the downward sloping demand curve.

In Figure 2, two demand curves are shown, corresponding to two different sets of investor valuation distributions. Both distributions feature the same mean expected return across investors, but in the case of the solid line, investors have a greater variance of expectations than in the case of the dashed line. The result is that when the vertical supply curve is drawn in at  $F$  shares, the market price is higher in the case of greater dispersion of opinions about the value of the stock. Even though the valuation is the same, greater dispersion of opinion leads to the steeper solid demand curve and the higher price,  $P^H$ . Because the fixed stock of shares ends up being owned by those who value it most highly, when opinions are very dispersed, the equilibrium price is driven up by the buying of those having extreme valuations.

A crucial point of Miller's model is that short selling is unlikely to drive an asset's price all the way to the mean valuation level. There are three reasons for this, all of them tending to limit the total amount of short selling that is likely to take place. The first is that there will likely be relatively few investors who think that shorting will be profitable. Only those investors who think that the asset's expected return will be negative will wish to short. Furthermore, our expectationally heterogeneous investors will only desire to short if they believe that they can make more (risk adjusted) money shorting one particular asset than they can going long in the assets about which they are most optimistic. In Figure 1, only those investors in the far end of the left tail, left of zero expected return, will consider short selling. Second, there is the risk, ever present in the Miller model, that a mispricing will widen rather than narrow. In terms of Figure 2, an arbitrageur who feels that the price  $P^L$  is too high may be discouraged from shorting the stock for fear that the distribution of investor valuations may become more extreme, increasing the slope of the demand curve from that of the dashed line to that of the solid line, thereby driving the price up from  $P^L$  to  $P^H$ . This "noise trader risk" will be discussed extensively below in the section on closed-end funds. Finally, there are institutional constraints that either ban short selling or make it very costly. The most obvious

are the outright prohibitions that preclude many mutual funds and institutional investors from shorting and the collateral and margin requirements that tie up capital when shorting.

For a retail investor, taking a short position means complying with the Federal Reserve's Regulation T, which requires retail investors to leave 150% of the value of the shares to be shorted as collateral to be held by their stock brokers. Until the position is covered, no interest is received on this collateral (see Duffie, Garleanu, and Pedersen (2002)). What this means is that a retail investor will only dare to short when he believes that the shorted stock will fall so far in value as to make up for the opportunity cost of the lost interest that could have been earned on his collateral. In terms of Figure 1, the number of short sellers would be even less than the area under the distribution curve to the left of zero expected returns. Only the extreme left part of the distribution tail would choose to short, as only those investors expecting substantially negative returns will feel that the short position will make up for the opportunity cost of interest forgone on the collateral.

Short selling is less costly for major investors such as hedge funds because they are not subject to Regulation T. They only have to put up 102% of the value of the shorted stocks as collateral. Furthermore, they also receive interest payments on the cash they put up as collateral. Because of this, major investors would presumably be more willing to take up short positions. But as documented by Geczy, Musto, and Reed (2002), the greater the potential gains to shorting, the less interest they receive on their collateral.

The collateral put up by short sellers is considered a type of loan made by the short seller to the share lender. Because the collateral is a loan, interest must be paid to the short seller putting up the collateral (i.e., making the loan). The interest rate that is paid is referred to as the "rebate rate" because the borrower in effect rebates to the lender some of the money he is earning by investing the collateral. Being a callable, short-term loan, the rebate rate is normally near the money market rate so that the short seller gets paid a fairly decent rate of

return on his collateral and is not subject to large opportunity costs while holding an open short position. But when a stock is in great demand by short sellers, those lending shares can negotiate to pay lower rebate rates. As a result, short sellers of stocks that are in great demand for shorting purposes end up receiving very little (and sometimes even negative) rates of return on their collateral. This mechanism serves to reduce the amount of short selling activity seen in the markets. Only those major investors who believe that a stock's price will fall enough to cover the opportunity costs of shorting will dare to short. And if there is great demand to borrow shares for shorting, those opportunity costs will rise.

The net effect of these restraints on short sellers is that the price of a stock will be driven down less far than it would be were short selling more convenient, involved less collateral, and was less costly in terms of rebate rates. This is illustrated in Figure 3, where  $F$  is again the total number of shares outstanding and  $S$  represents the number of shares sold short.  $F + S$  is therefore the total supply of shares sold onto the market. The increase in supply causes the market price to fall from  $P^0$  to  $P^1$ . But unless the supply increase due to shorting is substantial, the price will not fall all the way down to  $P^M$ , the price consistent with the mean expected valuation across investors.

A dynamic formal model incorporating heterogeneity of investor expectations and costly short selling has been developed by Duffie, Garleanu, and Pedersen (2002). It shows that not only can prices remain high in the presence of short selling, they can even rise higher than the valuation of the most optimistic investor because the market capitalizes in the future revenues that can be had by lending out shares to short sellers.<sup>3</sup> What is more, short sellers in the model are unable to drive prices down to the levels they consider proper. In the long run, the price-increasing valuations of optimistic investors keep prices high.

This formalization of the Miller (1977) insight will hopefully direct attention away from the CAPM model which fails so spectacularly in practice towards Miller's simple supply and

demand framework, which is capable of better explaining observed investor and asset pricing behavior. In particular, it must be noted that Miller's model can deftly account for the three famous violations of the CAPM model noted above.

First, we should expect a high amount of trading because investors have heterogenous beliefs about the future. As the valuation of the marginal investor in Figure 1 changes, so will the market price. Anything, no matter how irrational, that affects the distribution of investor expectations, will likely cause trading, as the finite supply of shares will be bid away by those investors having, at any moment, the most optimistic valuations regarding the asset.

Second, we should not expect all investors to buy the market portfolio because each investor will have individual beliefs about which stocks are likely to do well in the future. Each investor's portfolio will end up being filled with those stocks that each investor is most optimistic about. Unless, by chance, all investors have the same beliefs about all assets, they will *not* hold the same portfolio. Rather, they will tend to hold those stocks for which they are members of the optimistic, far right tail of the expectations distribution.

Third, we should not expect the return on an asset to vary proportionally with its correlation with the market portfolio. Under the Miller model, capital returns on individual assets will be determined by how much the changing valuation of the changing marginal investor changes the market price of the asset. This price variation need not be correlated with the return on the market portfolio. That is because the market return is itself merely the sum of the returns on the assets comprising the market portfolio, and each of those returns is itself caused by changing marginal investor valuations. There is no reason to believe that the changing valuations of one asset will necessarily be correlated with those of any other asset or with the sum of the changing valuations across all assets.

The greatest strength of the Miller (1977) model and its subsequent formalizations is its ability to explain why the most basic implication of rational asset pricing, the law of one price, is so often violated. The next section applies the Miller (1977) model to five well known violations.

### **III. Famous Violations of the Law of One Price**

Each of the most famous asset pricing anomalies is a violation of a no-arbitrage condition. In every instance, one can show that two assets with identical future payout streams have different current prices. This is, of course, not possible under CAPM. Because the rational investors of the CAPM model agree that the two payout streams are identical, they never even attempt to give different prices to the two assets. Put slightly differently, market segmentation does not matter in the CAPM world. If all investors are homogeneous in their expectations, then even if you divided them into two groups and only let members of the first bid on one identical asset, and members of the second bid on the other identical asset, both assets would be priced identically. By contrast, segmentation matters greatly in the presence of expectational heterogeneity.

To see this, divide a group of heterogeneous investors in half, into a pessimistic half and an optimistic half (e.g., the left and right halves of the symmetric distribution of Figure 1). If you then set the two groups to bidding in separate markets on identical assets, the equilibrium prices will be different, with the price among the pessimistic investors being less than the price among the optimistic investors.

There are many examples of assets with equal future payouts having different current prices. Several of these will be discussed shortly. Each instance can be explained by heterogeneous investors being segmented into bidding the two identical assets to different prices. What

is more, the observed price differences do not appear to be rapidly erased by arbitrage. That is, when two assets with identical payout streams are given different prices by segregated groups of investors with different levels of optimism or pessimism, arbitrageurs appear to be slow to force prices to parity. Two factors appear to account for this. The first is the reduction of short selling caused by outright bans on shorting, the large collateral requirements involved, and the difficulty and delay often involved in merely locating shortable shares. The second is that noise trader risk appears to greatly limit the amount of capital that arbitrageurs are willing to commit to either long or short positions, as will be discussed extensively below in the section on closed-end funds.

We now turn to five of the more famous “anomalies” of behavioral finance and explain how each of them is well explained by the Miller (1977) model of segmented, heterogeneous investors bidding for shares in the presence of short selling and noise trader risk.

### **A. ARCO, Exxon, and Prudhoe Bay**

Norman (1971) relates how the market price of Atlantic Richfield Company (ARCO) increased much more than that of Exxon after the discovery of the rich Alaskan oil field at Prudhoe Bay—despite the fact that the two firms had equal half interests in the field. This is a clear violation of the law of one price, but one that is consistent with segregated groups of investors having different levels of optimism about the expected returns from developing the Prudhoe Bay deposits.

According to Oswald (2001), the initial announcement of the discovery was made in an ARCO press release in March, 1968. Figure 4 shows that this is the month in which one sees ARCO’s share price begin to rise rapidly.<sup>4</sup> That same month, Exxon’s price rises only marginally. In subsequent months, their divergence is even more extreme. In April, Exxon’s

price actually falls while that of ARCO continues to rise. When a second ARCO press release announces in June that a second gusher at Prudhoe Bay confirms that the deposit is the largest ever in North America, ARCO's price continues to skyrocket so that between the end of February and the end of July its share price increases from \$98.375 to \$184.00. During that same period, Exxon's price only increases modestly, from \$67.75 to \$78.25.

While inconsistent with CAPM and asset pricing models with homogeneous agents, the divergent behavior of ARCO and Exxon is explainable under the Miller (1977) model. Simply put, different investors with different valuation distributions caused there to be different responses to the same information.

Indirect evidence that the investor pools of the two firms were likely substantially different before the discovery can be seen in Figure 5, which gives the monthly trading volumes of the two firms around the time of the press releases. During February of 1968, only 86,500 shares of ARCO were traded, compared to 592,100 shares of Exxon. This suggests that a smaller and likely different pool of investors traded in the two stocks. Over the next few months, trading of both issues increased substantially, but that of ARCO grew much faster. The June volume for ARCO was 421,600 shares traded, while that of Exxon was 1,342,500. In other words, trading in ARCO increased 387% while trading in Exxon increased just 127%. This by itself is further evidence that the investors trading in ARCO reacted very differently to the press releases than those trading in Exxon.

But bubbles burst. The price of ARCO collapsed in August, falling to \$91.25, despite the fact that during the same month price of Exxon rose slightly. One might see in this sudden collapse the action of arbitrageurs. But it is interesting to note that after August 1968, the trading volume levels of the two stocks were of the same order of magnitude whereas prior to March 1968, ARCO's trading volume was an order of magnitude smaller than that of Exxon.

As the bubble burst, the two share prices may have come to reflect the same valuation of the Prudhoe Bay discovery not because of the action of short sellers but because after several months of trading, the shares of the two companies came to be held by the same group of investors. This would have eliminated the segmentation that had previously existed between the owners of ARCO shares and those of Exxon shares. As this segmentation was eliminated, the same group of investors—and more importantly, the same marginal investor—would have priced the discovery equally, thereby eliminating the price difference that originally resulted because the two shareholder groups consisted of different people with different valuation distributions.

## **B. Royal Dutch Petroleum and Shell Oil**

A second famous asset pricing anomaly also comes from the oil industry. In 1907, the Dutch firm Royal Dutch Petroleum merged operations with the English firm Shell Transport and Trading Company LLC. As part of the merger, they agreed to split profits on a 60-40 basis. Because of this, their stocks should be priced at a similar ratio. Royal Dutch trades both in the Netherlands and on the NYSE where it is part of the Standard and Poor's 500 Index, while Shell trades in London and is part of the Financial Times Stock Exchange Index. Despite being heavily traded and highly liquid in both markets, Rosenthal and Young (1990) and Froot and Dabora (1999) find deviations of up to 35% away from the expected 60-40 ratio.<sup>5</sup>

What is more, the level of deviation seems not to be explainable in terms of fundamental factors such as exchange rate risk or differences in tax laws. Rather, Froot and Dabora (1999) find that the share prices of the two companies are highly correlated with the returns of the markets in which they respectively trade. When the US stock market does well, Royal Dutch shares do well; and when the UK stock market does well, Shell shares do well.<sup>6</sup> This is of

course not consistent with CAPM or rational asset pricing models, but is again easily explained using the Miller (1977) framework wherein the price of each company in each country is affected by the level of optimism of the marginal investor in each country. When the marginal US investor becomes more optimistic about share prices in general, the price of Royal Dutch rises along with the rest of the US market, and similarly with the marginal UK investor and Shell shares. Unless the marginal investors in the two countries happen to have similar outlooks across the full range of available assets, we should not expect that prices of the two firms will fall in the proper 60-40 ratio.

The deviation of Royal Dutch and Shell share prices away from the 60-40 ratio is often not only substantial but lingering. It has even defied the best attempts of arbitrageurs to narrow the gap. In 1998, the infamous hedge fund Long Term Capital Management had to unwind, at a loss, the \$2.3 billion position it had taken in Royal Dutch/Shell. The fund had attempted to profit by going long the shares of one firm and short the shares of the other, expecting the gap between them to diminish. As documented by Lowenstein (2000), the fund was forced to unwind the position as the gap widened rather than narrowed. The lesson to be learned from this is that changes in the opinion distribution and the valuation of the marginal investor can overcome the best efforts of short sellers to profit by forcing prices back to proper, arbitrage-free levels.

This particular type of risk is discussed by Shleifer and Vishny (1997), who point out that arbitrage activities only work if asset prices move in the “right” direction. In the Miller (1977) model, the movement of asset prices in such a “right” direction cannot be taken for granted as there is no sure way to predict how the distribution of heterogeneous investor beliefs is likely to evolve. And given a limited ability to undertake risk, due either to limited time horizons or limited capitalization, arbitrageurs may be forced to unwind positions at a loss just as they are becoming more potentially profitable. That is, Long Term Capital Management had to unwind

its Royal Dutch/Shell position when the gap between their share prices widened. But such an increase represents an even greater potential profit opportunity. Just as the opportunity got better, the fund had to liquidate.

In a world in which arbitrageurs face the risk that price deviations may widen rather than narrow, they will be less likely to commit capital towards undertaking the sort of arbitrage activities that could force prices to parity. Given this reluctance, we should not be surprised to see deviations of the Royal Dutch/Shell variety.

### **C. Palm, 3Com, and Equity Carve-outs**

Stark violations of the law of one price are sometimes observed when parent companies sell off subsidiaries. Such sales proceed in two steps. First, a “carve-out” takes place. This is an initial public offering at which the parent company sells a fraction of the shares of its soon-to-be independent subsidiary to the general public. Later, a “spin-off” happens. This is when the remaining shares of the subsidiary are given to the shareholders of the parent company. This is done at some pre-designated ratio so that for each single share of the parent company owned, an investor will receive  $X$  shares of the newly independent subsidiary.

The key point is that between the carve out and the spin-off, there are two ways of obtaining shares of the subsidiary. You can either buy them directly on the secondary market, where the carve-out shares are now trading. Or you can buy a share of the parent company, knowing that for each share of the parent company you will soon receive  $X$  shares of the subsidiary. If arbitrage pricing held, then the ratio of the prices of the parent and subsidiary firms stocks should be at least  $1:X$ , as this ratio equalizes the price of obtaining shares in the subsidiary through either direct purchase or purchasing shares of the parent company.

Lamont and Thaler (2001) report that this arbitrage-free pricing ratio was violated several times during the Tech Bubble of the late 1999's. The most famous example, reported upon at the time by the Wall Street Journal and the New York Times, was the case of 3Com Corporation's sale of its Palm Computing subsidiary. 3Com sold 5% of Palm to the public on March 2, 2000, at which time it also announced that it would spin-off the remaining shares to 3Com shareholders by the end of the year at a ratio of 1.5 shares of Palm for each share of 3Com owned. At the end of the first day of trading, Palm closed at \$95.06 per share. At a ratio of 1:1.5, that implied that each share of 3Com should cost at least \$142.59. Note that this imputed price is a lower bound, as it assumes that the rest of 3Com was worthless. Indeed, given that 3Com was a successful company, one would have expected it to trade for substantially more than \$142.59. However, the actual closing price of 3Com that day was only \$81.81. Not only is this far less than the minimum no-arbitrage price, it is so low as to be an example of a "negative stub value."

The stub value of a firm is the implied stand-alone value of the parent company once it spins off the remaining shares of its subsidiary. After the first day of Palm trading, the stub value of 3Com was negative. Given that corporations are limited-liability entities, it should be impossible for shares to have any price lower than zero. Therefore, the negative implied value of 3Com was a huge deviation from market rationality.

However, the Palm/3Com example is not unique. Lamont and Thaler (2001) find five other cases of negative stub values out of a sample of only 18 carve outs. The 18 were all instances where a parent firm had retained at least 80% of the subsidiary's shares at carve-out and had given written announcement that the parent company would spin off all of the remaining shares. The written announcement also normally declared that the spinoff of all shares would take place shortly, usually within 6 to 12 months.

These criteria are very important because they should have tended to reduce the sample to cases where arbitrage is more likely to take place, thereby reducing the likelihood and magnitude of negative sub values. By telling the markets that a full spinoff will be accomplished quickly, short sellers will have a more clear picture of the risks involved. In particular, they know that on the day that the spinoff takes place, arbitrage will in fact force prices to parity—meaning that their arbitrage position will pay off with certainty on the spinoff date. To see this, imagine an arbitrageur on March 2, 2000 buying a share of 3Com for the closing price of \$81.81. This will, by the end of the year, entitle him to 1.5 shares of Palm. It also entitles him to whatever the value of 3Com will be once those 1.5 shares of Palm are spun off. But this remainder, this stub value, will not be less than zero, as the price of 3Com cannot fall to less than zero. Given that the stub value on March 2, 2000 is negative, this means that you will be guaranteed a profit on the spin off day: the stub part of your investment must rise in value from something negative to at least zero. Done more elaborately, you buy one share of 3Com, short 1.5 shares of Palm, and wait until the spinoff day. When it comes, your long and short positions in Palm will cancel out, and you'll gain whatever the price of 3Com is after the spin off.

The sample chosen by Lamont and Thaler (2001) is therefore one in which we should be very unlikely to see negative stub values. Companies that say they will spin off all shares and give a short horizon for doing so guarantee arbitrageurs that they will be able to make a sure profit in a the course of just a few months. Consequently, it is quite remarkable that of the 18 firms, fully 6 have negative sub values. Because such a large fraction of carve outs have negative stub values even under conditions which are unfavorable to negative stub values, it is perhaps not surprising that Mitchell, Pulvino, and Stafford (2002) are able to find 82 cases of negative stub values in US equity markets over the period 1985-2000.

While the phenomenon of negative stub values is inconsistent with the law of one price and homogeneous investor expectations, it is fully consistent with Miller's framework. All that is required to explain negative stub values is for there to be a segmentation between the investors trading Palm and those trading 3Com and enough barriers and disincentives to prevent arbitrageurs from quickly forcing arbitrage-free prices.

As with Long Term Capital Management's failure to force arbitrage-free pricing in the case of Royal Dutch/Shell, arbitrageurs also failed to rapidly force arbitrage free prices in the case of Palm/3Com. Whereas Figlewski and Webb (1993) find that only 0.2% of the float of a typical company is shorted at any given time, huge short selling was undertaken in the case of Palm and the five other negative-stub companies studied by Lamont and Thaler (2001). In fact, short interest in Palm peaked at 147.6%, or almost 1.5 times the number of shares outstanding.

But despite such massive short selling, the 3Com stub remained negative for almost two months after short sales were initiated. This is even more striking when you consider that shorting cannot begin until 20 days after an IPO, at which time physical stock certificates are delivered to brokerage houses, which can then lend them out to short sellers. Given that the Wall Street Journal and the New York Times had published articles on 3Com's negative stub the day after the Palm IPO, investors were well aware of the profits to be made by shorting. It is amazing, then, that the negative stub persisted for almost two months after arbitrageurs had 20 days to prepare to take out short positions.

One can only conclude that there was a massively different distribution of investor valuations among Palm investors and 3Com investors. The difference was so great, in fact, that massive short selling could only partially rectify the two prices. The short selling did increase the supply of Palm and drive down its price, but the supply could not increase fast enough to quickly and fully offset the optimism of the marginal Palm investor.

## D. Violations of Put-Call Parity

Another way to see the inconsistent pricing of Palm and 3Com shares is to examine another arbitrage condition, put-call parity. To understand this condition, note that the following two portfolios will have the same payout on the expiration date,  $T$ , where today is considered to be time  $t = 0$ . The first portfolio consists of a European call option with a strike price of  $X$  as well as an amount of cash equal to  $Xe^{-rT}$ . The second portfolio consists of a European put option with the same strike price of  $X$  plus a single share of the underlying stock. The identical payout of both portfolios on the common option expiration date,  $T$ , will be  $\max(S_T, X)$ , where  $S_T$  is the price of the underlying stock on the expiration date,  $T$ . Because both portfolios will have the same payout on date  $T$ , they should trade today for the same price. If  $c$  is the current price of the call option,  $p$  is the current price of the put option, and  $S_0$  is the current price of the underlying stock, then, the following no-arbitrage condition, the put-call parity condition, should hold:

$$c + Xe^{-rt} = p + S_0. \quad (1)$$

Lamont and Thaler (2001) find that Palm options displayed massive violations of the put-call parity relationship of equation (1).<sup>7</sup> What is more, a slightly different no-arbitrage condition holds that for at-the-money put and call options, the calls should cost more than the puts.<sup>8</sup> This condition is also massively violated, with at-the-money Palm puts costing about twice as much as at-the-money Palm calls on March 17, 2002.<sup>9</sup>

Since puts give the right to sell in the future, it is clear that the marginal investor in the options market felt that price of Palm should be much lower than did his counterpart in the direct Palm shares market. Segmentation appears to have allowed assets with identical payouts

to have different prices in different markets. This is even more evident if you realize that one can “synthesize” a share of Palm by simultaneously buying a call, selling a put, and holding the present value of their common strike price. This portfolio will have the same value on the common expiration date of the two options as a share of Palm stock. Consequently, the cost of setting up synthetic long today should equal the present price of Palm stock. To see this, re-arrange equation (1) as,

$$c - p + Xe^{-rt} = S_0. \quad (2)$$

Lamont and Thaler (2001) find this condition massively violated for Palm and the other cases of negative stub values.<sup>10</sup> Depending upon the time left until expiration, the price of the synthetic long position in Palm was up to 23% less than the price of a share of Palm bought directly. This indicates that the investors involved in the market for Palm options believed that prices would be much lower on the relevant expiration date than did those investors trading Palm shares directly.

That the price of a synthetic long position in a stock can differ from that of the underlying asset is shown by Ofek, Richardson, and Whitelaw (2002) to be a wide-spread phenomenon that appears to be directly related to how difficult the given underlying asset is to short. Ofek, Richardson, and Whitelaw (2002) examine the options of all stocks trading in the USA over the period July 1999 to November 2001. They divide this period up into 118 dates that are approximately 5 trading days apart and then filter the data on these 118 days by examining only stocks that are non-dividend paying and which have intermediate-maturity pairs of at-the-money put and call options of the same expiration date. This gives them 80,614 pairs of options on 1734 stocks over the 118 weekly trading dates.

Their first significant finding is that violations of equation (2) are asymmetric. That is, you are much more likely to find in the data that  $c - p + Xe^{-rt} \leq S_0$  than you are to find that  $c - p + Xe^{-rt} \geq S_0$ . That is, if arbitrage is violated, it is much more likely that the synthetic long will cost less than the real stock rather than vice versa.

This finding makes sense from the perspective of segmented markets operating under short-selling constraints. If the price of a stock falls below the price of the synthetic long, it is easy for arbitrageurs to rectify the situation. They simply go out and buy shares of the stock, driving up the share price until it equals the price of the synthetic long. Matters are quite different if the price of the stock rises above the price of the synthetic long. In such instances, the mechanism that could act to equalize the prices is short selling. But, if short selling is hampered in any way, this pressure to move towards parity will be limited and convergence will be slow.

The second major finding of Ofek, Richardson, and Whitelaw (2002) is that as the difficulty of shorting a stock increases, the more likely the price of the underlying is to exceed the price of the synthetic long. This is evidence that the greater are the limitations on short-sale arbitrage activities, the more the prices of two identical assets are able to vary because of segmentation and differences in the valuation of the marginal investors trading the two identical assets.<sup>11</sup>

Ofek, Richardson, and Whitelaw (2002) utilize rebate rates on short sale collateral as their measure of the difficulty of shorting a stock. For stocks which are easy to short, the rebate rate is usually approximately equal the money market rate. However, if there is great demand by short sellers to borrow the shares of a particular company, then the lenders of such shares can negotiate to pay lower rebate rates on the short seller's collateral. Because there is no proper market for obtaining shares to short, observed rebate rates may not be equilibrium prices in the sense of equalizing the supply and demand for shares to short. Ofek, Richardson, and

Whitelaw (2002) argue, though, that their rebate rate data can serve as a good proxy for the difficulty and cost of obtaining shares to short. Consequently, one should expect to find it more likely that a stock's price will exceed the price of its synthetic long if rebate rates are lower (i.e., if lenders of shares can drive a hard bargain because of the high short seller demand for their shares.) To test this, Ofek, Richardson, and Whitelaw (2002) match the options pairs in their data set to a data set containing rebate rates. As expected, they find that the harder it is to short, as measured by lower rebate rates, the greater is the excess of the share price over the price of the synthetic long. In fact, an astonishing 12.23% of the 80,614 observations are cases where the price of the stock is greater than the price of the synthetic long, even after taking account of transactions costs. Failures of put-call parity are thus probably the most common and pervasive violations of the law of one price to be found in financial markets.

Synthetic long prices, however, can be greater than, as well as less than, the prices of actual longs. It is instructive to plot out a histogram of their relative prices. Let  $S$  denote the spot price of a stock, i.e. the price of the actual long. And let  $S^*$  denote the price of the synthetic long for that stock. If the two positions had equal prices, it would be the case that  $\ln(S/S^*)$  would equal zero.

In the authors' data set, there were 56,072 options pairs for which rebate rates were near the market rate of interest and therefore short-sales constraints appeared to be non-binding. There were an additional 8,699 options pairs for which rebate rates were significantly less than the market rate of interest and therefore short-sales constraints appeared to be strongly binding. Figure 6 plots out for each of these two groups a distribution histogram of their log price ratios,  $\ln(S/S^*)$ . The solid line gives the relative frequency distribution for the 56,072 options pairs for which short-sales constraints do not bind. It is narrow and centered on the no-arbitrage value of zero at which the synthetic long and actual long would have equal prices. By contrast, the dotted line that gives the distribution of the 8,699 options pairs for which short-

sales constraints are tightly binding is heavily skewed to the right, with a mode of 0.5, implying that  $S/S^* = 1.65$ , or that, at the mode, prices of actual longs are 65% higher than the prices of synthetic longs. Another way to measure the effect of binding short-sales constraints on distributional symmetry is to compare cumulative distributions. When short-sales constraints do not bind, 47.3% of  $\ln(S/S^*)$  observations are less than zero, whereas only 27.7% are less than zero when short-sales constraints do bind.

The asymmetric distribution of put-call parity violations suggests that they are well explained by the Miller (1977) model. One must assume only that the valuation distributions are different in the options markets and the stock markets. The asymmetry follows directly upon the imposition of short sale constraints, which make difficult arbitrage activities in cases where the stock costs more than the synthetic long, but which do not increase the difficulty of arbitrage in cases where the synthetic long costs more than the stock.

## **E. Closed-end Fund Discounts and Premia**

The role of short selling in constraining deviations between the prices of two assets with identical payouts is also important for closed-end funds, which are mutual funds whose shares trade like stock. Because the contents of their portfolios must, by law, be published weekly, it is possible for investors to exactly replicate the portfolios of closed-end funds, so that they could obtain the same payout stream either by buying the shares of a fund or by replicating its portfolio. Consequently, one would expect the market value of a fund's shares should to equal the market value of the fund's portfolio. This no-arbitrage condition is typically violated: Closed-end funds often trade at substantial discounts and premia to the value of their underlying portfolios.

To be precise about these discounts and premia, let  $N_t$  denote the net asset value per share of a given fund. The net asset value (NAV) is simply a fund's portfolio value less any liabilities the fund may have; it is the value of the fund. Let  $P_t$  be the current price of the fund's shares on the stock market. The discount or premium at which the fund's shares trade is defined to be  $D_t = P_t/N_t - 1$ . Values of  $D_t > 0$  indicate premia, while values of  $D_t < 0$  are referred to as discounts.

Figure 7 shows the discounts/premia at which the shares of the largest closed-end fund, Tri-Continental Corporation, traded over the period 1980 to 2001. The fund has ranged over the past two decades from trading at substantial discounts of over -25% to small premia of about 5%. Even more interesting, changes in Tri-Continental's discount/premium have often been very rapid, making them hard to explain in terms of changing transactions costs, changing tax laws, or other changing fundamentals. Liquidity is also not an issue. Tens of thousands of its shares trade each day on the NYSE. Portfolio replicability is also not a problem. Tri-Continental holds only large, liquid stocks in its portfolio.<sup>12</sup>

The discounts/premia of closed-end funds appear to present a further example of segmented markets pricing identical assets at different prices. But this is not to say that closed-end fund investors are totally unaware of fundamentals. In fact, they seem to rationally take into account the fact that funds should trade at a modest discount in order to capitalize out expected future management fees.<sup>13</sup> This can be seen in Figure 8, which plots, for 464 funds over the period 1985-2001, a relative frequency histogram of the 225,306 weekly discount and premia observations that fall into one-percent wide bins ranging from a discount of -50% to a premium of +50%.<sup>14</sup> As is clear from the diagram, the mode of the distribution is -6%, which is very close to the rational discount of -7.2% predicted by the model of Flynn (2002). This model takes into account management fees, stochastic fund death times, and the fact that discounts/premia should increase with fund dividend payout rates. The value of the model

is confirmed by the fact that these three fundamental factors collectively explain 56% of the cross-sectional variation in fund discounts/premia, as demonstrated by Flynn (2003).

Figure 8 also indicates that deviations of discounts/premia from the rational discount level of -7.2% are approximately Gaussian, with most of the weight of the distribution concentrated near the mean. Small deviations are common, but larger ones are progressively more rare. Many deviations, though, are too far from the mean to be consistent with rational pricing. In particular, fully 31% of the 224,306 weekly discount observations are of premia. Only investors who believed that fund managers could beat the market would have been willing to pay premia to buy into a fund rather than purchase its portfolio directly. In fact, it is the willingness of investors to buy funds at premia that allow them to come into existence. As related by Weiss (1989), all closed-end funds are priced at their IPOs at a 10% premium, this premium being necessary to raise the cash needed to pay the investment bankers for their IPO services.

Lee, Shleifer, and Thaler (1991) provide evidence that closed-end fund investors are different from the investors who directly buy the shares held by funds in their portfolios. In particular, the shares of closed-end funds are owned almost entirely by small investors while those held by funds in their portfolios are heavily concentrated in the hands of institutional investors like pension plans, insurance companies, and mutual funds. If the valuations of the marginal investors in these two groups are not equal, then we should expect, in the Miller (1977) fashion, that the prices set by the two groups will also be different. Only the action of arbitrage will help to equalize valuations and drive funds to trade at the rational discount level of about -7.2%.

As with the asymmetry of put-call parity violations, however, the inability to sell short causes there to be an asymmetry in deviations from the rational discount level. In particular, deviations towards large premia are more common than deviations towards large discounts.

This is not immediately apparent because the histogram of Figure 8 truncates the tails of the distribution. It is the case, though, that while the lowest observed discount from 1985-2001 was -66.5%, the highest premium was +205.4%.<sup>15</sup>

This asymmetry in deviations is likely due to the great difficulty to be had in shorting the shares of closed-end funds. D'Avolio (2002) examines the market for short sale stock lending by combining exchange data giving short interest on stocks with a proprietary loan data base obtained from one of the largest securities lenders in the world, a firm so large that, in each month over the examined 18-month period from April 2000 through September 2001, its outstanding loan balance was more than 10% of total market short interest. D'Avolio (2002) reports that 27% of closed-end funds appear not to have been shorted by anyone during that period, and that of the remainder that had been shorted, 92% were not reported at all in the loan data base.<sup>16</sup> This implies that 27% of closed-end funds may have been impossible to short by any means, and that of the remainder that were shorted by some investor, 92% could not have been easily shorted. This is because if large securities lenders do not have shares available, anyone wishing to short a closed-end fund must deal with the cumbersome process of asking his broker to run a "locate," whereby his broker calls around to various brokerage houses attempting to find out if they have in their inventories any shares available for shorting. As reported by Duffie, Garleanu, and Pedersen (2002), this process may take weeks and brokers often cannot locate enough shares for their clients, thereby having to give them only "partial fills."

The upshot is that the difficulty of shorting the shares of closed-end funds explains why extreme premia are more common than extreme discounts. If a fund is trading at an extreme discount, arbitrageurs can simply buy shares of the closed-end fund on the stock exchange, thereby driving up their price and reducing the magnitude of the discount. On the other hand,

if a fund is trading at a substantial premium, the path that arbitrageurs would like to take—shorting the fund—is often impossible or at least arduous and slow.

But given the ease of buying closed-end fund shares, why are there still so many observations of discounts below the median of -6%? Put differently, why is the histogram of Figure 8 basically symmetric, rather heavily skewed to the right? If the ability to buy fund shares makes the correction of discounts easy to rectify, why do we find so many of them? Their presence is likely the result of another factor which discourages arbitrage activities between two segmented groups even when such activities are convenient, quick, and low cost. This factor is noise trader risk, which was described and formally modelled by DeLong, Shleifer, Summers, and Waldmann (1990).

A noise trader is an irrational agent whose trading activities are inherently unpredictable. His trading can, however, affect asset prices. In particular, he can drive the price of Asset A away from that of the otherwise identical Asset B if he trades only in the market for Asset A but not in the market for Asset B, which is, instead, totally dominated by rational traders. Note that this setup is similar to the case of closed-end funds, where, as documented by Lee, Shleifer, and Thaler (1991), small (irrational) investors dominate the trading of closed-end fund shares while large (rational) institutional traders dominate the trading of the shares held in fund portfolios. Flynn (2002) argues that the noise trading in closed-end funds results from investors ever changing beliefs about the ability of fund managers to beat the market. When they are more optimistic, fund prices rise and  $D_t$  increases. When they grow more pessimistic, fund prices fall and  $D_t$  decreases.

The reason that the irrational fund investors are not fully offset by rational investors—even when the later have no problems at all taking either long or short positions—is because the noise traders create a non-diversifiable risk factor that rational traders must take account of. Specifically, suppose that a fund is trading at a very large discount of -35%. One might expect

rational traders to buy shares of the fund, as doing so would, at a lower cost, obtain the same future payout stream as would replicating the fund's portfolio. But taking such a position in the presence of noise traders is potentially quite dangerous. The noise traders might become more pessimistic about the fund, sell its shares and drive down the price of the fund, thereby causing a capital loss for the rational traders who though buying at a deep discount was a good idea. The risk of this sort of thing happening is not theoretical. It is precisely what happened to Long Term Capital Management when the price gap between Royal Dutch and Shell widened instead of narrowed. And it is an ever-present danger when investing in closed-end funds.

This can be seen by examining Figure 9, which takes from the set of all discount/premium observations graphed in Figure 8 the subset of discounts between -25% and -20% and sees how it evolves over time, again grouping discount/premium observations into 1%-wide bins to form histograms.<sup>17</sup> The top graph of Figure 9 gives the initial distribution, the middle graph the distribution one month later, and the bottom graph the distribution one year later. Any arbitrageur who hoped to make a safe profit by buying the shares of funds trading at deep discounts would likely have felt aggrieved because any such undertaking is in fact very risky. After 1 month, nearly as many of the initial discounts have widened as have narrowed, and many of them have changed substantially. And after a year the spread is even more extreme. What figure 9 demonstrates is that noise trader risk is real and quite substantial. And it must be taken into account by any arbitrageur wishing to make money off the difference between the price of a closed-end fund's shares and the cost of replicating its portfolio.

Figure 9 does not, however, give a full appreciation of the true volatility caused by noise traders. That is because it analyzes what happens to an initial group of *discounts* as time passes. Arbitrageurs wishing to take advantage of deep discounts can easily achieve their goal by buying fund shares in the stock market. As we have see, though, arbitrageurs wishing to take advantage of deep premia are hampered by the difficulties of short selling. Because

of this, we should expect to find that noise traders have more of a free reign when funds trade at premia. This is exactly what is shown in Figure 10, which gives the distributional evolution of all premia falling initially between +20% and +25%. To ease comparison, the scaling of the vertical axis of each of the three sub-graphs is the same in both Figures 9 and 10. What we see by comparing the two figures is that the rate of diffusion of discounts/premia away from initial levels is much faster for premia than discounts. Because it is so difficult to short, fewer arbitrageurs are in the markets when funds trade at premia. Without their stabilizing influence, fund prices are more volatile. Viewed slightly differently, a comparison of the graphs shows that noise trader risk is still substantial even under the volatility reducing influence of arbitrageurs. Arbitrageurs mitigate but do not eliminate the price volatility caused by noise traders.

Because rational traders must always fear noise trader risk, they will not take large positions attempting to drive fund prices to rational levels. They avoid doing so because closed-end funds do not actually offer true arbitrage opportunities, which by definition are riskless. Such riskless opportunities come about when one can simultaneously buy and sell identical assets at different prices. You simultaneously buy at the lower price and sell at the higher price. The key factor is the simultaneity. With a closed-end fund trading at a deep discount, you could simultaneously go long the fund and short the underlying portfolio, hoping to profit from a convergence in prices. But this is not the same as a simultaneous buy and sell. A long-short position in a closed-end fund hedges every risk except one, noise trader risk. Unless noise trader risk goes to zero—unless, that is, all the noise traders go away—risk averse arbitrageurs will not attempt to fully offset the mispricings caused by noise traders.

A profound consequence of the self-limitation of arbitrageurs in the presence of noise trader risk is that discounts/premia only very slowly mean revert. This is evident in Figure 11, which plots initial discounts/premia versus discounts/premia 52-weeks later, using one-

percent wide bins for initial discounts/premia. For instance, all discounts falling between -25% and -24% were identified and of this subset, the average and standard deviation 52-weeks later were computed. Because 98.1% of all discount/premium observations fall between -30% and +25%, ignore outliers by concentrating on the middle of Figure 11. What it clearly shows is that even after 52 weeks, there is only very weak mean reversion. This can be seen by comparing the bold average line with the line of dots, which gives what would happen if discounts/premia showed absolutely no mean reversion. Mean reversion is in fact so slow that fund weeks where the average discount was -20%, for instance, still had, on average, a discount of -16.5% after 52 weeks.<sup>18</sup>

It thus appears that in the face of noise trader risk, a risk which cannot be hedged because it is by definition uncorrelated with anything else, funds show only the most modest tendency to correct discrepancies between fund share prices and the value of fund portfolio holdings. This is an important finding because similar behavior may well occur in other markets.<sup>19</sup> If noise trader risk limits the willingness or arbitrageurs to restore rational pricing, then mispricings can persist indefinitely, or at least until the segmentation between the two identical assets can be removed.

As documented by Brauer (1984), closed-end funds sometimes vote to liquidate themselves or convert into open-end funds. Either case implies that shareholders will be able to liquidate their holdings at par with the value of the underlying portfolio. Brauer (1984) finds that as soon as such decisions are made public, any discount or premium disappears. This is clear proof that arbitrageurs are ever vigilant. The large discounts/premia seen on closed-end funds do not arise because arbitrageurs are unaware of the mispricing. They arise because noise trader risk discourages attempts to arbitrage the price difference. This point is also made by Mitchell, Pulvino, and Stafford (2002) in their study of negative stub values. They find that the average time between the first appearance of a negative stub value and its termination

is 236 days because the high risks involved discourage arbitrage activity. They indeed find that, net of short selling costs, returns to arbitrageurs are “just barely larger than the risk-free interest rate.”

## **IV. Finite Horizons and Arbitrage Limitations**

The distribution of arbitrage-pricing violations among closed-end funds is more or less symmetric, with there being nearly as many under-pricings as over-pricings. By contrast, most violations of the law of one price appear to be asymmetric. This is especially true of the put-call parity violations found by Ofek, Richardson, and Whitelaw (2002). There, the large majority of violations are of stocks trading for more than, rather than less than, the cost of constructing a synthetic long position. One is led to ask, Why are violations of the law of one price equally likely to be on the up side or the down side among closed-end funds while those of other assets only tend to be on the up side?

The reason that immediately comes to mind is that the inability to short gives rise to the large number of up side violations of put-call parity. But I think this misses the point. The deeper reason is that options contracts are of finite duration.

If a stock's price is below the cost of constructing a synthetic long, one can guarantee a profit in finite time by buying a share of stock and going short a synthetic long because the options involved have well-know, finite expiration dates. By contrast, closed-end funds are on-going companies. Each fund will eventually go out of business but there is no way to tell when.<sup>20</sup> Consequently, an arbitrageur who wishes to go long the shares of a fund trading at a discount while shorting its underlying portfolio has no idea how long it will take to realize a profit. He will eventually profit when the fund eventually either liquidates or converts to an open-end format because the fund's shares will then trade at par with the value of the portfolio.

But in the mean time, the discount can widen and cause him to have to put up more collateral or even close out his position at a loss, as described in detail by Shleifer and Vishny (1997). In the presence of noise traders, the chances of this happening are ever-present. And the danger of noise trader risk applies to both up side and down side violations of the law of one price.

By contrast, the nature of options contracts greatly mitigates these risks. Their finite time horizons mean that a profit can be made for certain in a certain amount of time. Given the finite time horizons, the difference in price between the share and its synthetic long is extremely close to being a pure, riskless arbitrage opportunity. Because puts and calls are in limitless supply (because investors may write either contract), there is never a problem acting on the profit opportunity arising when the share price is less than the price of a synthetic long. You simply buy shares and write synthetic longs. The problem, and the cause of the asymmetry, arises when the share price rises above the price of the synthetic long. You can easily buy more synthetic longs, but because shares are in finite supply, anything that makes them hard to short will limit your ability to execute the arbitrage. Because short selling is in fact hard in the real world, we are left with an asymmetric distribution of put-call parity violations.

Making it easier for arbitrageurs to short would eliminate the asymmetry by basically eliminating the right tail of a distribution that mostly has only a right tail. Relatively few violations of put call parity would remain, on either the up side or the down side. By contrast, making closed-end funds easier to short would only make the distribution of violations more symmetric. But it would be largely unchanged, as the uncertain time horizon of closed-end funds means that any arbitrageur must subject himself to noise trader risk for an unknown amount of time. In the face of such risk, arbitrage will be limited and noise traders, largely unchecked, will be able to move prices to either discounts or premia as their changing beliefs dictate.

## V. Conclusions

Violations of the law of one price are very common. While these violations are incompatible with the traditional paradigm of rational investors sharing homothetic expectations about asset returns, they are readily explained if you presume that investors with heterogenous expectations trading in segmented markets will tend to set differing prices unless arbitrage pressures are robust.

Empirically, however, arbitrage activities are often not robust enough to equalize prices across segmented markets. Their limited effectiveness has two principal causes. The first is short sales restrictions. These manifest themselves as outright prohibitions on shorting, as large opportunity costs arising from having to put up large amounts of collateral, and as the difficulty and delay often encountered when trying to locate shortable shares. The second limitation on arbitrage, noise trader risk, affects both long and short positions. Since it is by definition an idiosyncratic, non-diversifiable risk, it serves to discourage arbitrageurs from taking the sort of deep positions that would be necessary to equalize prices across otherwise segmented markets.

If arbitrageurs know that a violation of the law of one price will be eliminated with certainty in a finite time period, they will be much more willing to devote capital to eliminating the violation than if a violation is of uncertain duration. In the later case, parity between the asset prices could only be maintained by an arbitrageur if he were willing to invest the capital necessary to counter any deviational pressure, in essence acting as a price fixer standing ready to buy or sell as needed to maintain equal prices. Because noise trader risk discourages such activities regardless of the direction of the pricing violation, we find that a symmetric distribution of violations results when noise trader risk must be borne for an unknown length of time. By contrast, the distribution of violations in cases where violations will be resolved

with certainty in a finite length of time is asymmetric, the asymmetry caused by short-sale constraints which make it difficult to eliminate up-side violations. Were such short-sale constraints non-binding, the mis-pricings in such cases would be largely eliminated.

## References

- Bedi, Jaideep, and Paul Tennant, 2002, Dual-listed Companies, *Reserve Bank of Australia Bulletin* October, 1–13.
- Brauer, Gregory A., 1984, Open-ending Closed-end Funds, *Journal of Financial Economics* 13, 491–507.
- Chen, Joseph, Hong Harrison, and Joseph C. Stein, 2002, Breadth of Ownership and Stock Returns, *Journal of Financial Economics* 66.
- D’Avolio, Gene, 2002, The Market For Borrowing Stock, *Journal of Financial Economics* 66, 271–306.
- DeLong, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann, 1990, Noise Trader Risk in Financial Markets, *Journal of Political Economy* 98, 703–738.
- Dimson, Elroy, and Carolina Minio-Kozerski, 1999, Closed-end Funds: A survey, *Financial Markets, Institutions, and Instruments* 8.
- Duffie, J. Darrell, Nicolae Garleanu, and Lasse Heje Pedersen, 2002, Securities Lending, Shorting, and Pricing, *Journal of Financial Economics* 66.
- Fama, Eugene, and Kenneth French, 1992, The Cross-Section of Expected Stock Returns, *Journal of Finance* 47, 427–65.
- Figlewski, Stephen, and Gwendolyn Webb, 1993, Options, Short Sales, and Market Completeness, *Journal of Finance* 48, 761–777.
- Flynn, Sean Masaki, 2002, A Model of the Discounts on Closed-end Mutual Funds, the Quantification of Investor Sentiment, and the Inability of Arbitrage to Force Closed-end Fund Share Prices to Par, Ph.D. thesis University of California, Berkeley.
- Flynn, Sean Masaki, 2003, THE PORTFOLIO ALLOCATION EFFECTS OF INVESTOR SENTIMENT ABOUT THE ABILITY OF MANAGERS TO BEAT THE MARKET, *Under Review, Journal of Finance*.
- Froot, Kenneth A., and Emil M. Dabora, 1999, How are Stock Prices Affected by the Location of Trade?, *Journal of Financial Economics* 53, 189–216.
- Gallmeyer, Michael, and Burton Hollifield, 2002, An Examination of Heterogeneous Beliefs with a Short-sale Constraint, *Working Paper, Graduate School of Industrial Administration, Carnegie Mellon University* pp. 1–40.
- Geczy, Christopher C., David K. Musto, and Adam V. Reed, 2002, Stocks are Special Too: An Analysis of the Equity Lending Market, *Journal of Financial Economics* 66, 241–269.
- Harrison, J. Michael, and David M. Kreps, 1978, Speculative Investor Behavior in a Stock Market with Heterogeneous Expectations, *Quarterly Journal of Economics* 92, 323–336.
- Hull, John C., 2000, *Options, Futures, & Other Derivatives*. (Prentice Hall New Jersey).
- Jones, Charles M., and Owen A. Lamont, 2002, Short-sale Constraints and Stock Returns, *Journal of Financial Economics* 66, 207–239.
- Lamont, Owen A., and Richard H. Thaler, 2001, Can the Market Add and Subtract? Mispricing in Tech-stock Carve-outs, *NBER Work Papers* w8302, 1–60.
- Lee, Charles M., Andrei Shleifer, and Richard Thaler, 1991, Investor Sentiment and the Closed-end Fund Puzzle, *The Journal of Finance* 46, 75–109.

- Lintner, John, 1965, The Valuation of Risk Assets and the Valuation of Risky Investments in Stock Portfolios and Capital Budgets, *Review of Economics and Statistics* 47, 13–37.
- Lowenstein, Roger, 2000, *When Genius Failed: The Rise and Fall of Long-Term Capital Management*. (Random House New York).
- Miller, Edward M., 1977, Risk, Uncertainty, and Divergence of Opinion, *The Journal of Finance* 32, 1151–1168.
- Miller, Edward M., 2001, Why the Low Returns to Beta and Other Forms of Risk: Divergence of Opinion and Uncertainty-induced Bias, *The Journal of Portfolio Management* 27, 40–55.
- Mitchell, Mark, Todd Pulvino, and Erik Stafford, 2002, Limited Arbitrage in Equity Markets, *Journal of Finance* 57, 551–584.
- Morris, Stephen, 1996, Speculative Investor Behavior and Learning, *Quarterly Journal of Economics* 111, 1111–1133.
- Norman, Charles A., 1971, Economic Analysis of the Prudhoe Bay Oil Field, in M.A. Adleman, Paul G. Bradley, and Charles A. Norman, eds.: *Alaskan Oil: Costs and Supply* (Praeger Publishers, New York ).
- Ofek, Eli, Matthew Richardson, and Robert F. Whitelaw, 2002, Limited Arbitrage and Short Sales Restrictions: Evidence from the Options Markets, *NBER Working Papers* #9423.
- Oswald, Lori Jo, 2001, ARCO Strikes Oil at Prudhoe Bay, *Petroleum News Alaska* 6.
- Rosenthal, Leonard, and Colin Young, 1990, The seemingly anomalous price behavior of Royal Dutch/Shell and Unilever N.V./PLC, *Journal of Financial Economics* 26, 123–141.
- Scheinkman, Jose A., and Xiong Wei, 2002, Overconfidence, Short-Sale Constraints, and Bubbles, *Princeton University Working Paper* pp. 1–49.
- Sharpe, William F., 1964, Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk, *Journal of Finance* 19, 425–442.
- Shleifer, Andrei, and Robert W. Vishny, 1997, The Limits of Arbitrage, *Journal of Finance* 52, 35–55.
- Weiss, Kathleen, 1989, The Post-offering Price Performance of Closed-end Funds, *Financial Management* 18, 57–67.

## VI. Notes

<sup>1</sup>See, for example, Miller (2001).

<sup>2</sup>This is a key point demonstrating that the results of the model are not driven by any group having only a vague idea about the beliefs of other investors. In the model, you know when others disagree with you and by how much.

<sup>3</sup>This intuition that stock prices can be, in markets with heterogeneous investors and short-sale constraints, above the present value of expected payouts of even the most optimistic investor has been developed by Harrison and Kreps (1978), Morris (1996), and Scheinkman and Wei (2002).

<sup>4</sup>The share price and volume data for ARCO and Exxon are from the monthly file of CRSP, the Center of Research in Securities Prices at the University of Chicago.

<sup>5</sup>Deviations of a similar magnitude have been found for other dual-listed firms. See Froot and Dabora (1999), Rosenthal and Young (1990), and Bedi and Tennant (2002).

<sup>6</sup>This covariation of internationally dual-listed shares with local stock markets is confirmed by Bedi and Tennant (2002) for several Australian/British companies whose respective parts are listed in Sydney and London.

<sup>7</sup>Equation (1) holds exactly only for European options, which can be executed only on their expiration dates. Lamont and Thaler (2001) actually examine American options, which can be executed any time before expiration. Adjustments can be made, however, to account for this difference. *Mutatis mutandis*, put-call parity is still massively violated by Palm shares.

<sup>8</sup>The condition is  $S_0 - X \leq C - P \leq S_0 - Xe^{-rT}$ , where  $C$  and  $P$  are the current prices of the American call and put options, respectively. Be sure to assume that the strike price,  $X$ , equals the current price of the underlying,  $S_0$ . See pp. 178 of Hull (2000).

<sup>9</sup>See Table 6 of Lamont and Thaler (2001).

<sup>10</sup>Actually, they test the analogous, but slightly different condition which holds for American options. The only difference is that of an early exercise premium, as American options can, unlike European options, be exercised before the expiration date.

<sup>11</sup>Further evidence that equities are more likely to be over-priced when short sales constraints bind is provided by Jones and Lamont (2002) who find that from 1926 to 1933, shares with binding short-sales constraints have high valuations and low subsequent returns.

<sup>12</sup>For a recent comprehensive survey of the large closed-end fund literature, see Dimson and Minio-Kozerski (1999).

<sup>13</sup> Management fees are paid as a percentage of fund NAV, typically at the rate of 1% per annum. This is very large, especially when one considers that funds typically earn less than 10% per annum on their portfolios. Capitalizing future management fees sums to a large present value. Footnote 20 shows that funds appear to die off exponentially, at the rate of 3.64% per annum. This implies a half-life for a fund of about 30 years. One can combine published fee rates with the exponential death model to give a precise estimate of the expected present value of future management fees. Subtracting these from the NAV of a fund would cause it to trade at about a -7.2% discount.

<sup>14</sup>The data used to construct this histogram comes from Weisenberger/Thompson Financial's FundEdge data set, which contains time series data on each of the 464 closed-end funds that traded in the USA and Canada in 2001. Some of these funds were founded in the 1920's, others much more recently. So how far back I have time series data varies by fund. Also, the data set does not contain funds which went out of business prior to 2001. There is therefore the possibility of some sort of survivor bias. This should not matter for the issues discussed in this paper, however, as the pricing behavior documented here appears to apply to all funds as long as they remain active.

<sup>15</sup>In fact, the positive skewness of the distribution means that its mean, at -4.3%, is higher than its mode.

<sup>16</sup>To put these figures in comparison, note that only 1.8% of S&P 500 companies were never shorted during this period, and of those that had been shorted, less than 1% were not in the loan database. D'Avolio (2002) defines a stock to have been shorted if the amount of short interest exceeds  $\min(\$10,000, 0.01\%)$  of shares outstanding.

<sup>17</sup>Note that the vertical axis is re-scaled for each of the three histograms.

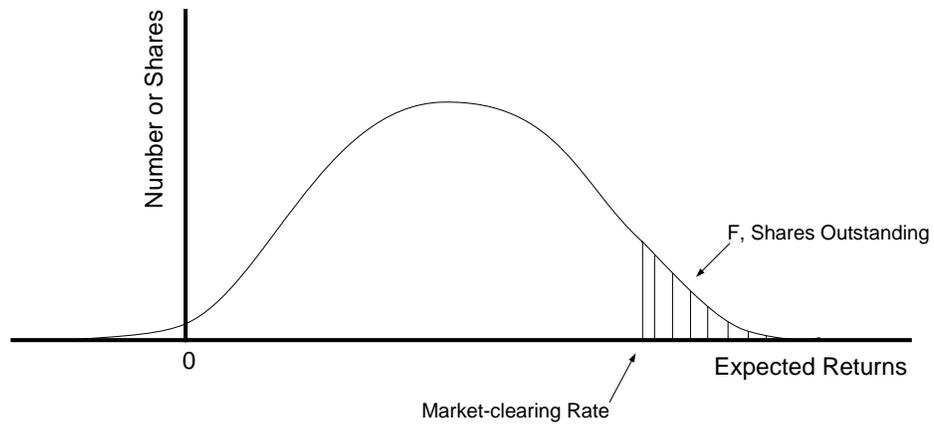
<sup>18</sup>Besides indicating that the reversion of discounts/premia to fundamental levels is quite slow, this figure also indicates that the market is aware of the correct level to which mean reversion should take place. This can be seen by noting that the line of dots and the bold average line cross at -6%. After a year, discounts of -6% do not go anywhere on average, while discounts/premia at other levels converge to this value. The slowness of mean reversion can also be quantified by running an AR(1) regression of current values of the discount/premium on those 1 year prior. Assume that discounts/premia are mean reverting to the level  $\bar{D}$ . Then they should follow  $D_{t+1} = \bar{D} + \phi(D_t - \bar{D}) + \varepsilon_t$ , where  $\phi$  gives the fraction of the period  $t$  deviation that remains the next year, and  $\varepsilon_t$  is a Gaussian shock. We can get empirical estimates for  $\phi$  and  $\bar{D}$  by running the regression,  $D_{t+1} = \text{constant} + \phi D_t + \varepsilon_t$ . Our estimated constant will be equal to  $\bar{D}/(1 - \phi)$ . Using monthly discount/premium data for the 464 closed-end funds in our sample, and estimating the equation using pooled least squares on data covering 1985-2001 gives a constant of -2.19 and an estimated value for  $\phi$  of 0.64. Using these estimates, we can back out an estimated value for  $\bar{D}$  of -6.08%. That is, discounts/premia mean-revert to a discount of -6%, but are so slow doing so that 0.64 of any deviation from the mean remains after one year.

<sup>19</sup>To test the noisiness of noise trader risk, I ran Fama French regressions for the 464 funds in my data set (see Fama and French (1992)). The dependent variable was the return from the hedge portfolio consisting of going long the shares of a fund and short its underlying. The average R-squared is a low 0.07, indicating that noise trader risk is truly an independent source of risk that arbitrageurs will not find easy to hedge. By way of comparison, the average R-squared from regressions where the only independent variable was the discount/premium

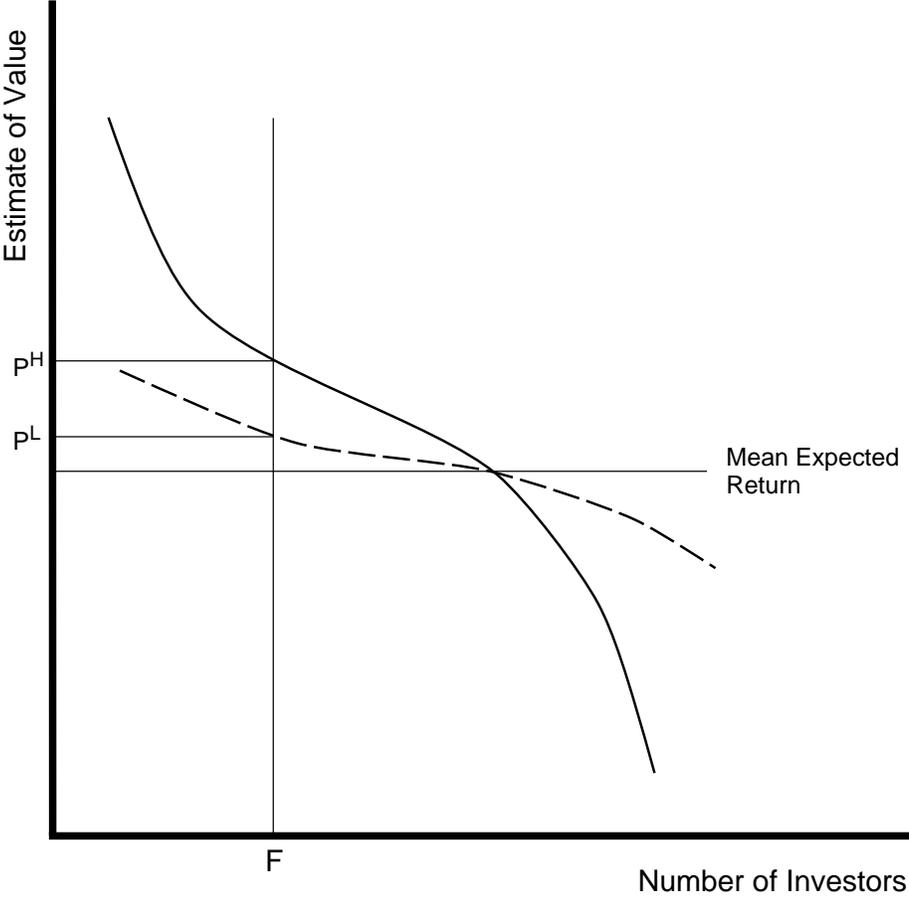
was 0.10. The systematic tendency for discounts/premia to mean revert is a better predictor than the Fama-French factors.

<sup>20</sup> Regressions of fund death rates on macro variables, market variables, and fund characteristic variables reveal no relationship with the probability of a fund going out of business. Fund death probabilities also appear to be independent of the age of the fund because they appear to follow a Bernoulli process. An examination of all closed-end stock funds trading on the NYSE from 1960 to 1999 reveals that one cannot reject the hypothesis that closed-end funds die off in a Bernoulli fashion, with an annual death probability of  $\gamma = .0364$ . Let  $X_t$  be the number of funds alive at the start of year  $t$  and  $O_t$  be the number of those that die during year  $t$ . Assuming that fund deaths are Bernoulli, with death rate  $\gamma$ , the expected number of deaths in year  $t$  is  $\gamma X_t$ . A Pearson's Chi-squared test statistic can therefore be constructed as  $D^2 = \sum_{t=1960}^{1999} \frac{(O_t - \gamma X_t)^2}{\gamma X_t}$ .  $D^2$  is distributed approximately  $\chi^2$  with  $1999 - 1960 - 1 = 39$  degrees of freedom. Our estimated  $D^2$  is 30.52 which is significantly less than than the 90% critical value of 51.81. We fail, therefore, to reject the hypothesis that fund deaths follow a Bernoulli process.

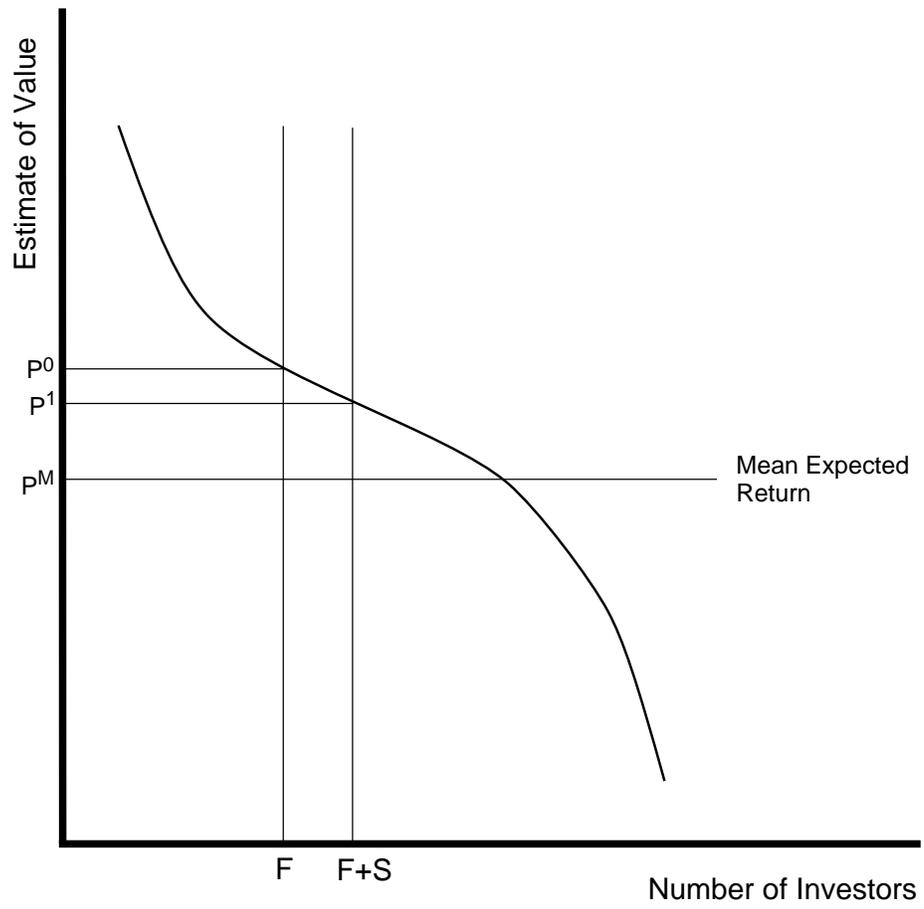
**Figure 1.** Distribution of Demand Due to Heterogeneity About Expected Returns



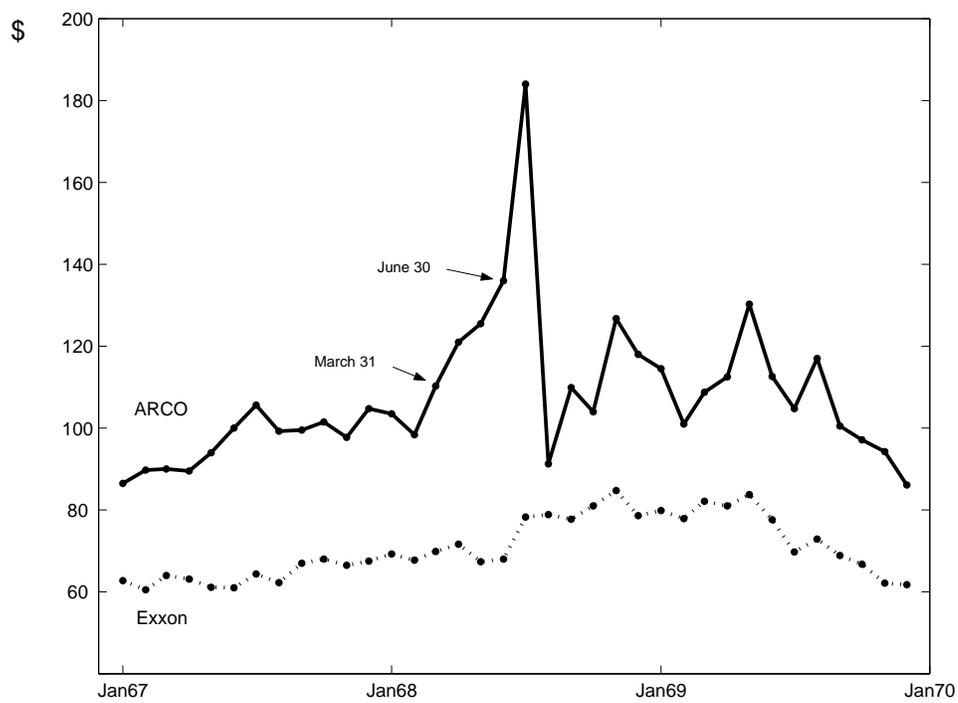
**Figure 2.** Two Asset Demand Curves, Each Resulting from Heterogeneity of Value Estimates



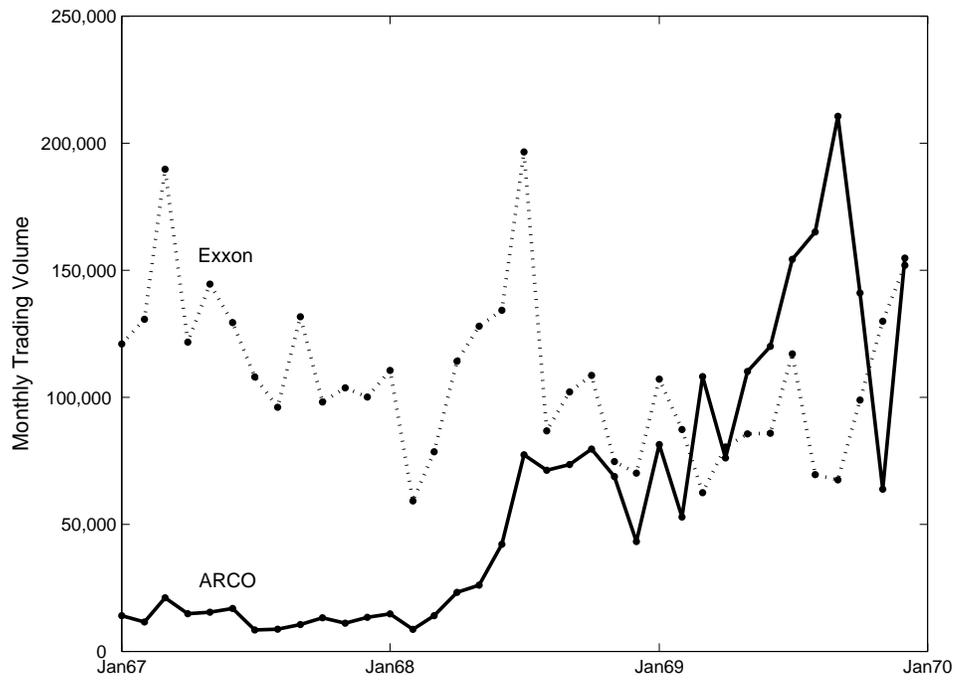
**Figure 3.** Decrease in Asset Price Due to Short Selling



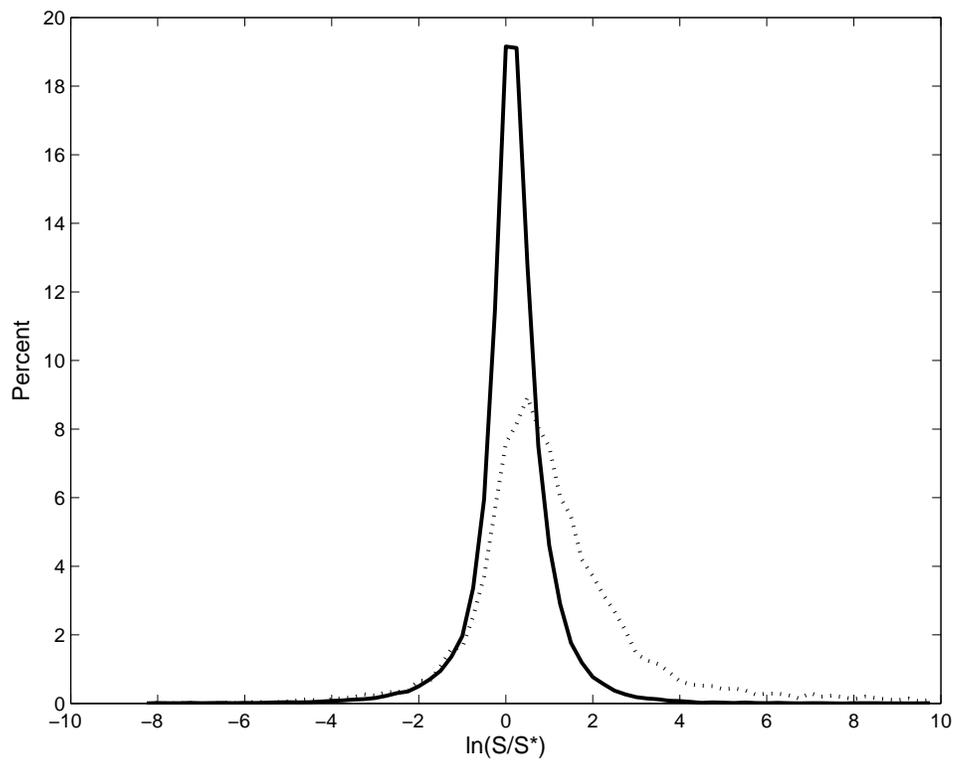
**Figure 4.** ARCO and Exxon Share Prices at Time of Prudhoe Bay Discovery



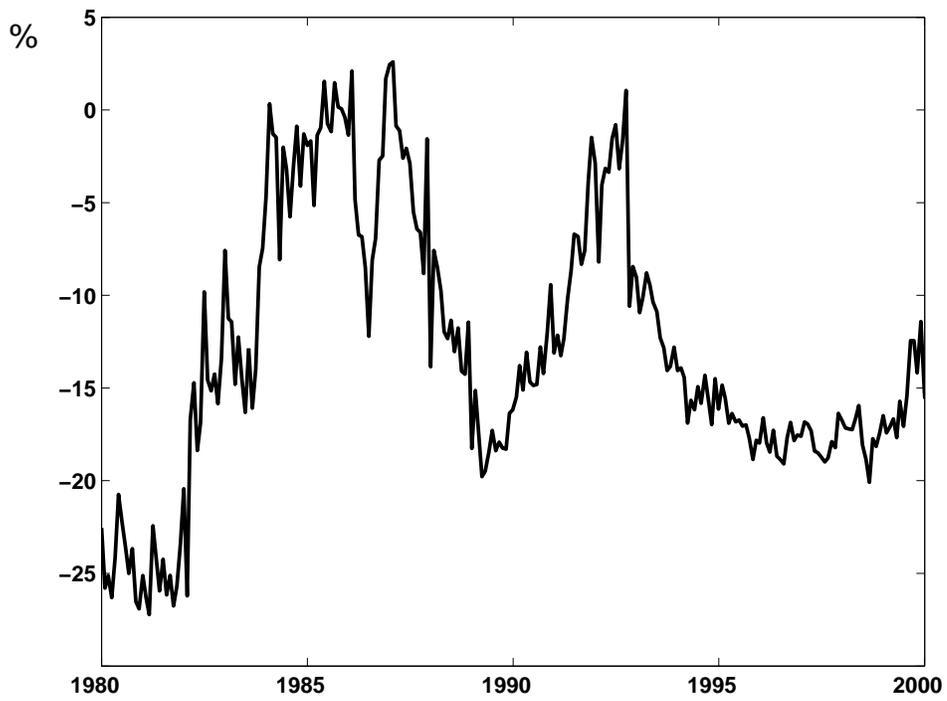
**Figure 5.** ARCO and Exxon Monthly Trading Volume at Time of Prudhoe Bay Discovery



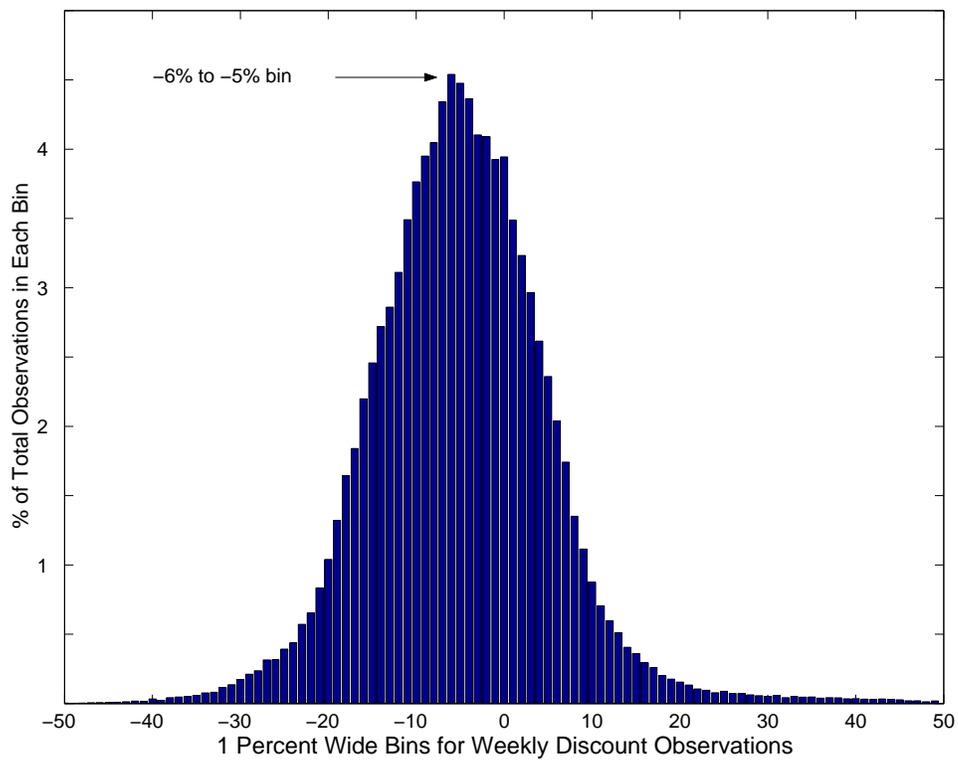
**Figure 6.** Relative Frequency Histogram of  $\ln(S/S^*)$  for the 56,072 Pairs where Short Sales Constraints were not Binding and for the 8,699 Pairs where Short Sales Constraints were Strongly Binding



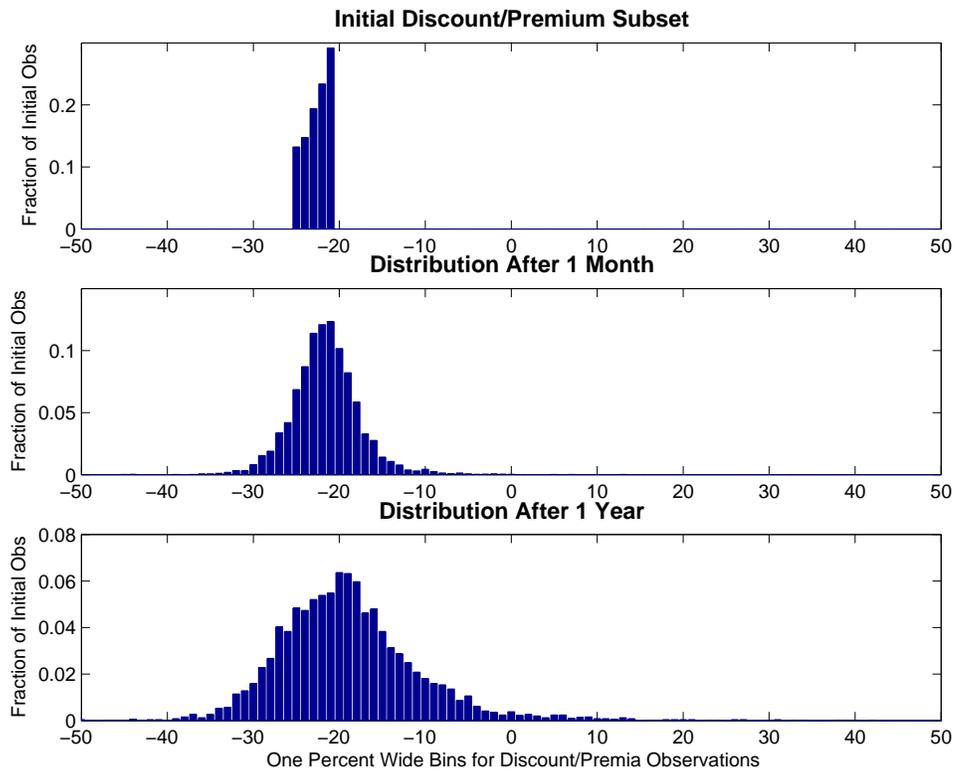
**Figure 7.** Tri-Continental Corporation Discount/Premium 1980-2000



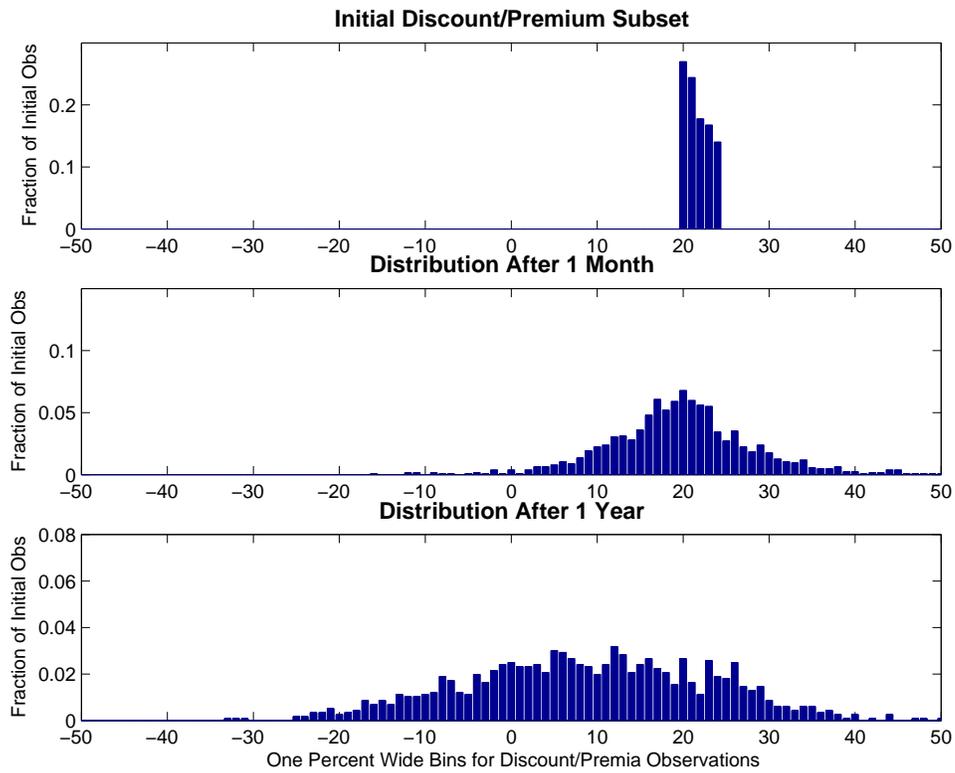
**Figure 8.** Distribution of Weekly Discounts/Premia, 1985-2001



**Figure 9.** Initial, 1-month Later, and 1-year Later Distributions of all 6,021 Discount Observations Between -25% and -20%



**Figure 10.** Initial, 1-month Later, and 1-year Later Distributions of all 1,250 Premium Observations Between +20% and +25%



**Figure 11.** Initial Discounts/Premia vs. Average and Standard Deviation 52 Weeks Later

