Available online at www.sciencedirect.com
ScienceDirect
Journal of Accounting and Economics 42 (2006) 35–52
JOURNAL OF Accounting & Economics
www.elsevier.com/locate/jae

Costly arbitrage and the myth of idiosyncratic risk ☆

Jeffrey Pontiff*

Carroll School of Management, Boston College, Chestnut Hill, MA 02467, USA

Available online 23 May 2006

Abstract

Transaction and holding costs make arbitrage costly. Mispricing exists to the extent that arbitrage costs prevent rational traders from fully eliminating inefficiencies. Although the relation between mispricing and transaction costs is well-known, the relation between mispricing and holding costs is misunderstood. One holding cost, idiosyncratic risk, is particularly misunderstood. Various myths are debunked, including the common myth that idiosyncratic risk matters because arbitrageurs only have access to a small number of projects [Shleifer and Vishny, 1997. The limits of arbitrage. The Journal of Finance 52, 35–55.]. The literature demonstrates that idiosyncratic risk is the single largest cost faced by arbitrageurs.

JEL classification: G00; G14; M40

Keywords: Capital markets; Costly arbitrage; Idiosyncratic risk; Market efficiency

1. Introduction

This paper strives to debunk seven myths surrounding the impact of rational arbitrage on irrational prices.

Myth 1: Arbitrage costs make arbitrage a zero-profit exercise.
Myth 2: Arbitrage costs decimate all arbitrage activity.

☆ Malcolm Baker, Pierluigi Balduzzi, Onur Bayar, David Chapman, Ken French, David Hirshleifer, Gang Hu, Cliff Holderness, S.P. Kothari (the editor), David McLean, Harold Mulherin, Shivaram Rajgopal, Philip Strahan, and Sheridan Titman have provided feedback on a previous draft. Gratitude goes to Denning Aaris who provided helpful copy editing. Special thanks go to Edward Rice for comments on an early draft, as well as a decade of numerous conversations about the practical implications of risk.
*Tel: +1 617 552 6786; fax: +1 617 552 8738.
E-mail address: pontiff@bc.edu.

0165-4101/$ - see front matter
doi:10.1016/j.jacceco.2006.04.002

Myth 3: Arbitrageurs avoid shorting high dividend paying stocks, since the short seller must rebate all dividend payments, making a short position more expensive.
Myth 4: Arbitrageurs care about idiosyncratic risk since they only have access to a small number of projects (Shleifer and Vishny, 1997).
Myth 5: Idiosyncratic risk has no impact on mispricing. Even if short-sellers are dissuaded by idiosyncratic risk, idiosyncratic risk will not affect the selling decision of sophisticated investors who already own the stock.
Myth 6: Since arbitrage costs are related to mispricing, the costly arbitrage literature provides proof that markets are efficient.
Myth 7: Costly arbitrage tests are subject to Fama's (1976) joint hypothesis to the same extent as classic tests of market efficiency.

Arbitrage is the transaction where a rational agent tries to profit from mispricing. Mispricing continues in equilibrium since costs prevent arbitrageurs from fully offsetting the price impact of irrational investors. The relation between arbitrage costs and market inefficiencies is often misunderstood. One arbitrage cost, idiosyncratic risk is particularly misunderstood. This misunderstanding likely stems from the detailed treatment of diversification and the capital asset pricing model (CAPM) in elementary finance text books that espouse that idiosyncratic risk does not matter. This logic is often falsely extended to make erroneous claims that idiosyncratic risk should not affect arbitrage activity, since the arbitrageur can costlessly diversify away this risk. Although the text book discussion of idiosyncratic risk is correct in terms of the CAPM, this logic does not carry over to arbitrage and market inefficiencies. Regardless of an arbitrageur's access to a diversified portfolio, and regardless of the number of available arbitrage projects, the idiosyncratic risk of each project will mute the arbitrageur's position in the project.

This article strives to clarify how inefficiencies may exist in equilibrium, and to illuminate the critical role played by idiosyncratic risk. This framework is useful for tests of market efficiency, since the interplay between idiosyncratic risk (as well as other arbitrage costs) and mispricing provides a testable prediction that only relies on the economic rationality of arbitrageurs, and not on the arbitrary behavior of noise traders. As will be shown in Section 5, the empirical studies that have explored such tests have affirmed the importance of idiosyncratic risk as the foremost arbitrage cost.

Recent research on inefficiencies and costly arbitrage has failed to emphasize the importance of idiosyncratic risk. This lack of coverage is attributable to two factors. First, some researchers believe variants of myth 4, and thus they downplay the importance of idiosyncratic risk as only a concern for investors with limited opportunities. Second, theoretical and empirical keystone papers have focused on systematic risk to the exclusion of idiosyncratic risk. One such theoretical paper, De long et al. (1990) argues that noise trader risk creates systematic risk, and asset prices are discounted to reflect this risk. Since there is only one asset in their model and all noise-traders have the same irrational expectations, idiosyncratic risk plays no role. The literature spawned by their paper has focused on irrational comovement, but has forsaken idiosyncratic risk.

Early keystone empirical work by Shiller (1989) tests whether stock index returns are too volatile relative to dividend streams. Since stock indices are not subject to idiosyncratic risk, the excess volatility calculation omits a large proportion of volatility. Pontiff (1997) estimates that 85% of closed-end fund excess volatility is idiosyncratic to a four-factor model that includes the Fama–French factors as well as the Lee et al. (1991) investor sentiment index. If this evidence is indicative of stock return excess volatility, stock return excess volatility is 6.7 times larger than implied by a Shiller-type test.

1.1. Costly arbitrage and market efficiency

The term "market efficiency" describes a market where information is immediately reflected in prices. The appeal of this view of a market follows from one of two assumptions. First, as is common in economics, agents are rational. Rationality implies that no one would buy an overpriced asset and that no one would sell an under-priced asset. Thus, the equilibrium of all rational agents implies price equates to fair value, and all information is reflected in asset prices. Put another way, in this equilibrium there are no dollar bills on the floor since rational agents do not drop bills on the floor.

A second assumption that is sufficient to generate an efficient market is that rational agents are able to profitably trade against mispricing until it disappears. The profit from these trades cease when price equals fair value, at which point the market is efficient (Friedman, 1953). In this equilibrium, there are no dollar bills on the floor, since all lost bills have been collected by profit-minded rational agents. For simplicity, we will call these rational agents "arbitrageurs."

Most financial economists would agree that markets are very efficient but not perfectly efficient. Thus, there may be some (but not too many) dollar bills on the floor, which begs the question, "what is wrong with our notion of a rational equilibrium?". The answer to the question starts with the acknowledgement that not all agents are perfectly rational when it comes to financial decisions. This follows from a large body of accumulated evidence that some people make financial mistakes. For example, Barber and Odean (2000) find that individuals tend to trade positions too often. Odean (1998) shows that investors tend to realize losses later, and gains earlier than tax-optimizing would suggest. Agnew (2005) shows that employees do not fully participate in 401(*k*)'s, they tend to over invest in company stock, and they follow irrational allocation heuristics.

Irrational decisions by some investors will not generate inefficiencies if profit-motivated investors counter their price impact. The price impact from the profit-motivated will be constrained to the extent that these traders are subject to costs, including the cost of idiosyncratic risk, that prevent them from fully taking advantage of the inefficiency. In this case, some dollar bills are left on the floor because they are too costly to pick up. This thought process leads to the notion of "costly arbitrage." The idea of constrained rational investors goes back, at least to Shiller (1984), who models interaction between irrational and rational traders. Shiller (1984) models an effective cost by preventing rational traders from devoting more capital to their trades than they have. The costly arbitrage framework detailed here follows Pontiff (1996), which considers how various arbitrage costs affect arbitrageurs' abilities to trade against inefficiencies. I use this framework to argue that closed-end fund anomalies (for example, Thompson, 1978, Pontiff, 1995, 1997) are the manifestation of mispricing.

2. Arbitrage costs and mispricing

Two costs prevent rational traders from exerting price pressure that totally eliminates mispricing: transaction costs and holding costs. Transaction costs are incurred when a transaction occurs. Examples of transaction costs are brokerage fees, commissions, and

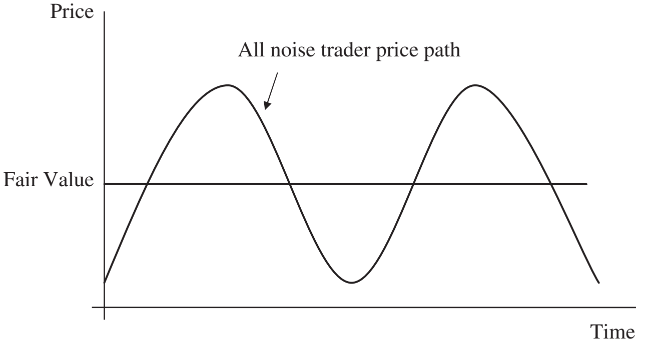

Fig. 1. Hypothetical mispricing process in the absence of arbitrageurs.

market impact. Holding costs are costs that are incurred every period that a position remains open. Examples of holding costs include the opportunity cost of capital, the opportunity cost of not receiving full interest on short-sale proceeds, and idiosyncratic risk exposure. Idiosyncratic risk exposure is a holding cost, but total risk is not a holding cost since risk that is related to the returns of other assets can be offset with hedge positions in those assets. Dividend payments are negatively related to holding costs for two reasons. First, dividends serve to decrease the effective duration of a position, and thus decrease the number of periods that the position is held. Second, holding costs are typically proportional to the value of the position. If a dividend is paid, it effectively lowers the level of the future stream of holding costs.[1]

Both transaction costs and holding costs have similar relations with mispricing: the greater the cost, the greater the potential mispricing. Costs create an impediment that decrease the ability of rational traders to trade against the mispricing.

Fig. 1 graphs potential mispricing relative to intrinsic value over time. Wave-like mispricing is assumed for illustrative purposes. This figure assumes that no rational traders are present. In the absence of costs, the presence of rational traders would force the wave into a straight, horizontal line at the value of the stock. In this case, there would be no mispricing. The price wave represents the all noise trader equilibrium.

Fig. 2 illustrates the equilibrium when transaction costs are the only costs faced by rational traders. Transaction costs have received predominant attention in studies of mispricing. Using the framework of Garman and Ohlson (1981), the two horizontal lines at x and $-x$ represent transaction costs bounds.[2] Within these bounds arbitrage is unprofitable. Outside these bounds arbitrage is profitable. Transaction costs alone create the unrealistic equilibrium of no price pressure within the bounds, and infinite price pressure outside the bounds, confining equilibrium mispricing within the bounds.[3] The

[1]This proposition is formally illustrated in the appendix.

[2]In a pure transaction cost equilibrium, the transaction cost bounds are determined by rational traders with the lowest transaction costs.

[3]Fig. 2 shows that the equilibrium price stays flat when price hits the bounds, which is consistent with the assumption that the irrational expectations of noise traders are unaffected by the arbitrage price pressure. Even if noise trader behavior is affected by arbitrage pressure, causing the price to revert within the bound, the implication of this figure is unchanged.

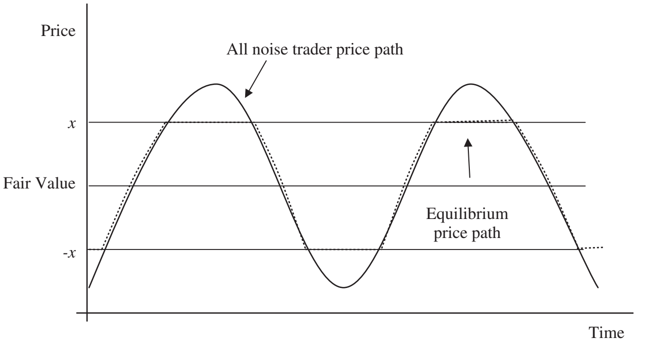

Fig. 2. Hypothetical mispricing process in the presence of arbitrageurs who only face transaction costs. x determines the size of the no-profit arbitrage bound.

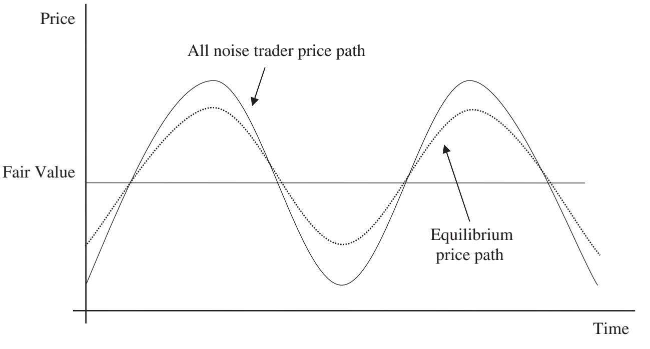

Fig. 3. Hypothetical mispricing process in the presence of arbitrageurs who only face holding costs.

implication for costly arbitrage is straightforward—the bigger the transaction costs, the bigger the potential for mispricing. Most of the literature on mispricing and arbitrage costs focuses on the transaction cost subset of arbitrage costs. This partial view of costly arbitrage is unsatisfying, in that it predicts arbitrage pressure is either non-existent or infinite.

Holding costs make even seemingly riskless arbitrage risky (Tuckman and Vila, 1992). A rational investor knows with certainty that a stock is overpriced by $10. In a world with no holding costs (and no transaction costs), the investor short sells the stock and waits until the price falls to fair value to cover his short position. In a world with a $1 per period holding cost, the arbitrageur will profit only if the mispricing dissipates within 10 periods. If the mispricing takes more than 10 periods to dissipate, the arbitrageur loses money. Fig. 3 illustrates the equilibrium when holding costs are the only costs faced by rational traders.[4] As long as price deviates from fair value, arbitrageurs always take a position.

[4]This figure ignores some potential foibles that may complicate the impact of rational traders, and thus it is better suited for a one period equilibrium than a dynamic problem. For example, arbitrage pressure may be path dependent as the wealth of arbitrageur changes or the arbitrageur may be a professional investor and subject to client-induced cash inflows and outflows (Shleifer and Vishny, 1997).

Despite this, they never take positions that are large enough to remove all mispricing. Arbitrageurs must trade-off the risk of their position against the expected profit of the position. Although arbitrage activity is profitable on average, the size of the position will be chosen such that the marginal profit of the arbitrage position is zero.

Myth 1: Arbitrage costs make arbitrage a zero-profit exercise.
Myth 2: Arbitrage costs decimate all arbitrage activity.

Both of these myths can be debunked as long as there are holding costs and a finite number of arbitrageurs with finite wealth. Arbitrage with holding costs is analogous to a free dinner buffet. A free buffet is a benefit to the recipient, yet the recipient does not endlessly eat. Rather, he or she eats only until the marginal benefit of eating equates to the marginal discomfort. Similarly, a mispricing opportunity is a benefit to the arbitrageur, yet he does not take an infinite position in the mispriced security. The mispricing equilibrium with holding costs is a richer description of reality than the transaction cost equilibrium—arbitrage occurs frequently, yet mispricing continues. In equilibrium, arbitrageurs and mispricing co-exist.

Myth 3: Arbitrageurs avoid shorting high dividend paying stocks, since the short seller must rebate all dividend payments, making a short position more expensive.

This myth ignores the fact that prices drop to reflect dividend payments. As long as an arbitrage position is subject to holding costs, arbitrageurs will do the exact opposite of what this myth suggests—they favor high dividend yield stocks, regardless of whether or not the mispriced security is over- or under-priced. Dividend payments effectively lower the duration of the arbitrage position, and thus, lower the expected holding costs from the position. As the Appendix A shows, this result can be demonstrated under very general assumptions.

On the surface, the fact that arbitrageurs prefer high-dividend paying stocks may seem to contradict the findings of Dechow et al. (2001), who show that high dividend paying stocks are less likely to be shorted. This interpretation of their findings ignores the possibility of a confounding effect; noise traders may be less likely to induce upwards price pressure on high dividend paying stocks.

The impact of dividends on mispricing deserves special consideration since dividend policy is determined by the firm. Thus, dividend policy provides a mechanism for management to influence the degree to which the firm is mispriced. Dividend initiation announcements are associated with positive stock price reactions (Asquith and Mullins, 1983), and positive post-announcement long-run returns (Michaely et al., 1995). These findings are consistent with the possibility that under-priced firms are more likely to initiate dividends. Arbitrage buying pressure increases upon dividend initiation announcement since the expected holding cost associated with the position have decreased, prompting an immediate positive price reaction. Arbitrage remains costly, thus some mispricing remains, as is evidenced by the post-announcement returns.

Dividend policy should have the opposite effect on overpriced firms. Jensen (2005) argues that overvalued share prices create incentives for management to destroy long-run value, and that corporate governance policy has been ineffective. A corporate commitment to pay dividends is one mechanism that should mitigate this problem, since dividends lower the costs to those who short sell the stock, inducing price pressure towards fundamental value.

As stated earlier, one holding cost, idiosyncratic risk, deserves special attention. This attention is warranted since idiosyncratic risk is the key holding cost associated with equity trading. Despite this, idiosyncratic risk is subject to many misconceptions. We will use two tools to think about idiosyncratic risk. One is based on a clinical study of a decision, and the second is based on the mathematics that is used to create optimal portfolios. Both of these investigations consider environments where stocks are mispriced relative to a broad index. This approach is used for ease of illustration. The case of a mispriced broad portfolio yields similar implications (Fama and French, 2005). A more general treatment that considers the temporary case of substantial mispricing that permeates broad factors is outlined by Daniel et al. (2001).

3. The role of idiosyncratic risk: the parable of the Sahara free ace coupon

The importance of idiosyncratic risk is well illustrated by an opportunity presented to me and a finance professor friend, Edward Rice, while we were on a trip to Las Vegas. Upon arriving, each of us acquired a "free ace" blackjack coupon at the Sahara hotel. The player uses the coupon by presenting it with their bet to the dealer before the cards are dealt. The dealer replaces the first card for the coupon user with an ace. This imparts a large advantage on the player. For the follower of the optimal blackjack strategy, this coupon implies a profit of 41.5¢ for each dollar wagered. Since the bet occurs over an instant, it involves no time value of money. In the jargon of corporate finance, the bet is a positive 41.5% net present value (NPV) project. The NPV of this project scales linearly, in that every incremental dollar devoted to the wager provides an incremental NPV of 41.5¢. Table 1 presents the approximate payouts and probabilities associated with the coupon.[5]

Given our understanding of the probabilities and cashflows associated with the coupon, Professor Rice and I had to make a decision regarding the amount of capital that we would devote to this gamble. Each coupon could only be used once, so we had to determine the total amount of money that we would gamble when we used the coupon. I recounted this story to hundreds of people, and have asked them, if they were in Las Vegas, at the Sahara casino (and thus, there was no transaction cost with using the coupon), how much would they gamble.[6] No one has ever answered that they would wager all of their wealth. What keeps people from gambling everything on an apparent positive NPV project? The user of the coupon is subject to idiosyncratic risk. If the risk associated with the hand of blackjack were not idiosyncratic, if it were correlated to other risks in financial markets, the coupon holder would create a riskless hedge with financial securities and bet an infinite amount. Unlike a financial decision made by a corporation that is owned by well-diversified shareholders, the free ace coupon holder can not diversify away the idiosyncratic risk associated with this project. Analogous to the dinner at a free buffet or an arbitrage opportunity, the coupon is valuable, but using the coupon induces a cost.

[5]In order to increase the simplicity of this example, I ignore the possibilities of splits, and doubling.

[6]Some neo-Puritan academics attach a stigma to gambling and are reluctant to directly answer this question as worded. In this situation, using the same probabilities and payoffs, I rephrase the question as "How much would you invest in an oil well, given the following payoffs and probabilities associated with the investment."

Table 1
Probabilities and cashflow outcomes associated with a one dollar wager with the Sahara free ace coupon

Probability (%)	Description	Financial outcome
31	Blackjack	1.5
25.5	Win	1
13	Match	0
30.5	Lose	−1
Expected cashflow $= 0.31 \times 1.5 + 0.255 \times 1 - 0.305 = 0.415$		

> Myth 4: Arbitrageurs care about idiosyncratic risk since they only have access to a small number of projects (Shleifer and Vishny, 1997).

At the center of myth 4 is the belief that access to only a small number of projects, and thus, under-diversification induces concern about idiosyncratic risk. This myth was first perpetrated by Shleifer and Vishny (1997), when they claim,

> 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.

They continue in saying,

> In reality, arbitrage resources are heavily concentrated in hands of few investors that are highly specialized in trading a few assets, and are far from diversified. As a result, these investors care about total risk, and not just systematic risk.

Although Shleifer and Vishny (1997) focus on the agency problems of professional arbitrage, their analysis of idiosyncratic risk has crept into the more general costly arbitrage literature. Almost every costly arbitrage published paper following Shleifer and Vishny (1997) has perpetuated this myth. Wurgler and Zhuravskaya (2002) in defending the use of idiosyncratic risk as an arbitrage cost, claim that idiosyncratic is relevant because, "First, as a practical matter, arbitrageurs are often not fully diversified .... Second, as a theoretical matter, there is simply no opportunity to diversify across S&P additions or similar events that occur infrequently through time." Ali et al. (2003) in a study of book-to-market ratios and idiosyncratic risk, make the same error when they claim, "... since arbitrageurs are not well diversified, idiosyncratic risk adds to total portfolio volatility ..." On a similar note, Mendenhall (2004) in a study of idiosyncratic risk and post-earnings announcement drift, declares, "Clearly, idiosyncratic risk should not be important for diversified investors." I will show, regardless of whether arbitrageurs have access to many or few arbitrage opportunities, idiosyncratic risk is clearly important.

The Sahara free ace coupon example illustrates the fallacy that permeates this myth. People who say that they would wager $200 with the free ace coupon, when asked how much they would wager if they had two coupons, usually say that they would wager about $200 per coupon. Thus, access to a second coupon does not affect the decision to use the first coupon. One conceivable reason why the second bet may be affected by the outcome of the first bet would be if the person demonstrates increasing relative risk aversion, such that they are willing to gamble more if their wealth has gone up from winning the previous gamble. The myth that having access to multiple coupons allows one to diversify risk, like the myth that having access to multiple inefficient securities allows one to diversify risk, is analogous to the time diversification myth that was debunked by Samuelson (1971). Diversification only works when risk is shared across agents (for example, Bodie et al., 2004). Having access to multiple coupons, as well as having access to multiple inefficient securities, does not involve sharing of risks. Sharing involves agreements between parties. Along these lines, before Professor Rice and I used our coupons we agreed to self-insure each other's losses (or gains). Because of this agreement, we intended to wager twice the amount we would have otherwise.[7] For the interested reader, it should be noted that our careful consideration was futile: upon arriving at the Sahara casino, we were informed by a blackjack dealer that the coupon only entitled us to a maximum wager of \$10.

4. The role of idiosyncratic risk: example with portfolio optimization

Another way to explicitly consider the impact of idiosyncratic risk is to revisit portfolio theory (Markowitz, 1952). Consider an investor with access to the risk-free security, the market, and n mispriced securities. In order to simplify this example, stock returns are generated from a one-factor model for which all covariance is attributable to market risk. Our analysis will assume that the investor is a price-taker, although the implications will hold in a more general model [such as Petajisto, 2004 or Fama and French, 2005].

A sophisticated investor can "repackage" each of the n mispriced securities into a hedge portfolio that holds the stock and shorts the appropriate amount of the market, such that each hedge portfolio has no covariance with the market. Thus, each hedge portfolio is only exposed to idiosyncratic risk. In an environment where covariance is determined by more than just the market factor, the hedge portfolio will instead contain short positions in all the relevant factors that are needed to reduce the position's volatility to pure idiosyncratic risk. This framework acknowledges that the stocks may be mispriced, not the market or other relevant factors. Thus, the alpha from the hedge position can be attributable entirely to the stock.

The return of the hedge position can be written as, r_i, where

$$r_i = \alpha_i + r_f + e_i. \tag{1}$$

Since the hedge position contains no market risk, if it is accurately priced, its return will equate to the risk-free rate, r_f. If the hedge position is mispriced, its alpha is denoted by α_i. e_i is the unexpected noise in the stock return. Since the position is fully hedged, its risk is independent to the noise in all the positions, and the total risk of the investor's portfolio variance, σ_p^2, may be written as

$$\sigma_\mathrm{p}^2 = \sigma_\mathrm{m}^2 x_\mathrm{m}^2 + \sum_{i=1}^{n} \sigma_i^2 x_i^2, \tag{2}$$

where x_i is the weight in security i. σ_i is the standard deviation of the hedge position—which represents only idiosyncratic risk, and σ_m is the standard deviation of the market

[7]The decision of two people to self-insure should lead to each person wagering the total amount that they would wager if they had two coupons. This occurs since the cashflow implication of using two coupons is identical to the cashflow implications of two people jointly using two coupons and splitting all winnings and losses.

return. Since all of the positions are uncorrelated, no covariance terms are included in Eq. (2).

Assuming a mean-variance investor, with a risk aversion parameter equal to λ, utility is defined as

$$U = x_f r_f + x_{\rm m} E(r_{\rm m}) + \sum_{i=1}^{n} (\alpha_i + r_f) x_i - \frac{\lambda}{2} \sigma_{\rm p}^2. \tag{3}$$

The optimal portfolio weights are solved by setting the first-order condition to zero, and recognizing that the individual weights must sum to one.

$$x_{\rm m} = \left(\frac{E(r_{\rm m}) - r_f}{\lambda \sigma_{\rm m}^2} \right), \tag{4}$$

$$x_i = \left(\frac{\alpha_i}{\lambda \sigma_i^2} \right). \tag{5}$$

Myth 4 (again): Arbitrageurs care about idiosyncratic risk since they only have access to a small number of projects (Shleifer and Vishny, 1997).

The hedge positions are designed to be only subject to idiosyncratic risk. Thus, they are uncorrelated to the market and the other hedge positions. Since the hedge positions are only subject to idiosyncratic risk, Eqs. (4) and (5) imply that the optimal exposure to the market is independent of arbitrage opportunities, and the optimal exposure to a specific hedge position is independent of all other risky positions in the arbitrageur's portfolio. If an arbitrageur discovers a new opportunity, the optimal weight is determined by his risk aversion, the alpha of the position, and the position's idiosyncratic risk. Other equity positions do not influence the amount of capital devoted to the arbitrage. Since exposure to the market portfolio is independent of arbitrage activity, capital for arbitrage is raised by either borrowing more, or decreasing the amount invested in the risk-free asset.

The falseness of myth 4 does not rely on the assumption that the arbitrageur can borrow. If anything, restricting borrowing makes myth 4 even more absurd, since new arbitrage opportunities must be financed by decreasing the amount of the portfolio that was dedicated to other arbitrage positions. Also, this myth is not specific to the mean-variance utility function, rather, as is shown by Pratt and Zeckhauser (1987) it holds for all standard risk-averse utility functions that are considered "proper," including exponential, power, logarithmic, and hyperbolic absolute risk aversion. For many of these utility functions the main result is even stronger—access to a new arbitrage opportunity *decreases* the optimal amount of capital devoted to other arbitrage positions.

The intuition behind Eq. (5) does not change if we relax the assumption that the covariance between the projects is zero. In the most general case, the optimal portfolio weight of an asset can be written as

$$x_i = \frac{\alpha_i}{\lambda \sigma_{e_i}^2}. \tag{6}$$

As Stevens (1998) shows, α_i is the intercept from a regression equation where the excess return of security i is regressed on the excess returns of all other risky assets. $\sigma_{e_i}^2$ is the idiosyncratic variance from this regression. Essentially, the difference between (5) and (6) is that (5) assumes that a hedged arbitrage position can be created by shorting the market, whereas (6) assumes that the optimal hedge is constructed from the universe of risky assets.

The preceding discussion focused on the arbitrageur's demand for mispriced securities, but it did not consider the arbitrageurs' price impact. In this sense, the results were the outcome of a partial equilibrium. An extension of the analysis to a full equilibrium still yields the result that idiosyncratic risk dampens arbitrage activity. For example, both Merton (1987) and Petajisto (2004) generate equilibrium models with informed and uninformed investors. Both papers derive rational investor positions that are remarkably similar to Eqs. (5) and (6), in that the optimal position size increases in alpha and decreases in idiosyncratic risk.

> *Myth 5*: Idiosyncratic risk has no impact on mispricing. Even if short-sellers are dissuaded by idiosyncratic risk, idiosyncratic risk will not affect the selling decision of sophisticated investors who already own the stock.

This myth relies on the false intuition that the idiosyncratic risk of a stock is only relevant to the extent that the sophisticated trader intends to hold the stock in his portfolio. As Eqs. (4) and (5) show, optimal portfolio weights are proportional to the expected alpha of the security divided by the security's idiosyncratic variance. A sophisticated trader who owns a stock and decides to sell it, does so because he has revised downward his estimate of the security's alpha. The magnitude of the original position of this stock was affected by the stock's idiosyncratic risk. Thus, ceterus paribus, a stock with higher idiosyncratic risk will have less selling pressure from sophisticated traders, since these traders have a smaller position than they would have if the stock had lower idiosyncratic risk.

5. Testing costly arbitrage

Most of the finance literature has given idiosyncratic risk little attention as an arbitrage cost. For example, Schwert (2004) in a review of anomalies mentions arbitrage costs in general, but not idiosyncratic risk (or holding costs) per se. Barberis and Thaler (2003) in a review of the behavioral finance and limits to arbitrage literature, acknowledge the theoretical importance of idiosyncratic risk, but only devote one paragraph to review an empirical example of idiosyncratic risk being costly (Wurgler and Zhuravskaya, 2002). The bulk of their theoretical and empirical discussion focuses on irrationality affecting systematic risk, either through market returns, returns to style portfolios, closed-end fund discount comovements, and comovement after index inclusions. Similarly, a survey article by Hirshleifer (2001) does a very competent job illustrating the theoretical impact of idiosyncratic risk, although he neglects to provide evidence of idiosyncratic risk's role.

Although many papers have focused on the relation between transaction costs and mispricing, this section will focus on the handful of papers that have studied the more interesting effect of holding costs and, in particular, idiosyncratic risk. Although this literature is still in its infancy, a common theme emerges—idiosyncratic risk is the single largest barrier to arbitrage.

Pontiff (1996) uses the framework of costly arbitrage to test whether closed-end fund discounts are a manifestation of mispricing. I posit that if closed-end fund discounts represent mispricing, then the absolute value of the discount will be positively related to arbitrage costs. Using natural log of the market capitalization of the fund as a proxy for the transaction costs of trading the fund, I find that larger funds tend to have smaller price discrepancies with net asset value. Regarding holding costs, the idiosyncratic risk of the fund's portfolio is strongly positively related to the absolute value of premium levels, as well as interest rates, a measure of the opportunity cost of capital. Consistent with the theory that dividend payments either lower the expected holding period of the arbitrage, or lower the future extent of the position, I show that dividend yields are negatively related to mispricing. The sum of my evidence implies that closed-end funds are subject to mispricing, and idiosyncratic risk is the predominate arbitrage cost that prevents the mispricing from dissipating.

Ackert and Tian (2000) examine mispricing of Standard and Poor's Depository Receipts (SPDRs) and MidCap SPDRs. Arbitrage in these products does not involve idiosyncratic risk since a trader knows the portfolio weights of the stocks that comprise the index, and thus can construct a perfect hedge. Besides idiosyncratic risk, arbitrage in these contracts involve little exposure to holding costs since the arbitrageur has the ability to exchange SPDR shares for the individual securities in the index, and vice versa. Thus, the arbitrage position can be liquidated at will.[8] As one would expect, SPDR mispricing is miniscule compared with closed-end fund mispricing. The authors never find mispricing greater than 1%. They also find no evidence of SPDR excess volatility; despite the fact that Pontiff (1997) finds that the typical closed-end fund is 64% more volatile than its portfolio. Ackert and Tian (2000) also compare mispricing between the SPDRs and the MidCap SPDRs. The stocks in the MidCap index tend to have higher transaction costs and higher dividends, thus one would expect their total arbitrage costs to be higher. Consistent with this, they find that MidCap SPDR mispricing is on average over nine times higher than SPDR mispricing.

Shleifer (1986) has shown that inclusion of a stock into the S&P 500 index is associated with an approximate return of 3%. Since inclusion to the S&P reveals little, if any, information about the firm, Shleifer interprets this finding as evidence of a downward sloping demand curve. Wurgler and Zhuravskaya (2002) extend Shleifer's study by investigating how the reaction is affected by the stock's idiosyncratic risk relative to the S&P 500. They find that stocks with lower idiosyncratic risk have a smaller reaction to inclusion, consistent with costly arbitrage activity.

Baker and Savasogul (2002) estimate abnormal returns from merger arbitrage, in which an investor takes a long position in a target and takes a short position in the acquirer. Although Baker and Savasogul do not explicitly compute idiosyncratic risk, they show that the takeover premium and the probability of takeover success contribute to the idiosyncratic risk of the long-short position. Consistent with idiosyncratic risk impeding price pressure to merger arbitrage positions, they find that a positive association between abnormal returns and their idiosyncratic risk proxies.

Ali et al. (2003) examine the relation between the book-to-market effect and idiosyncratic risk. They find that the ability of book-to-market ratios to predict future cross-sectional stock returns is stronger for stocks with more idiosyncratic risk relative to the CRSP value-weighted index. Their results are consistent with the book-to-market effect reflecting an inefficiency that arbitrageurs trade against. They find that when both

[8]The transfer agent charges a fee to exchange the SPDR, so the optimal strategy may sometimes involve liquidating the position in the open market.

idiosyncratic risk and firm size are included in their estimation, firm size is actually positively related to the book-to-market effect, although idiosyncratic volatility remains in the predicted direction.

Pontiff and Schill (2004) use a costly arbitrage framework to test whether long-run seasoned equity offerings (SEO) returns reflect mispricing. They find that SEOs with higher idiosyncratic risk relative to the Standard & Poor's 500 have more negative long-run SEO returns, consistent with idiosyncratic risk imposing a holding cost on sophisticated investors. They find that the relation between idiosyncratic risk and SEO returns renders an insignificant relation between SEO abnormal returns and firm size.

Ben-David and Roulstone (2005) investigate the abnormal returns of insider stock sales and purchases for semi-annual and annual holding periods. Consistent with costly arbitrage, they show that firms with higher idiosyncratic risk have more extreme abnormal returns. Their results do not compare idiosyncratic risk with other arbitrage costs.

Mendenhall (2004) studies how arbitrage costs affect post-earnings announcement drift. He shows that post-earnings announcement drift over the next quarter is stronger for firms with higher idiosyncratic risk, measured relative to the return of the S&P 500. He finds mixed results regarding the impact of transaction costs on abnormal returns. His most compelling proxy for transaction cost is trading volume, which is negatively related to drift, and consistent with the hypothesis that high-volume stocks have lower transaction costs. For all specifications, the *t*-statistics on idiosyncratic risk are much more significant than the *t*-statistics on volume.

Asquith and Muelbroek (1996) find evidence that short-selling interest predicts low cross-sectional stock returns. Duan et al. (2005) apply the reasoning of costly arbitrage to argue that if short-sellers are informed, then shorted stocks with high idiosyncratic risk stocks should experience lower abnormal returns than shorted stocks with low idiosyncratic risk. Consistent with idiosyncratic risk impeding informed investors, they find that high-idiosyncratic short-interest stocks have annual abnormal returns that are 13.5% lower than low-idiosyncratic short-interest stocks.

Mashruwala et al. (2006) show that firms with higher arbitrage costs tend to have higher exposure to the accrual anomaly. Although they find that transaction cost proxies are also associated with the anomaly, idiosyncratic risk tends to be the strongest variable. They also argue that their proxy for idiosyncratic risk, which is calculated with monthly returns relative to the CRSP equal-weighted index, captures the same variation as measures relative to the value-weighted index, and measures that are constructed with daily data.

A common theme that unifies this literature is that the primary source of arbitrage costs occurs from holding costs, and in particular, idiosyncratic risk. All of the above papers that simultaneously estimate the impact of idiosyncratic risk and transaction costs on mispricing, find that the impact of idiosyncratic risk on mispricing dwarfs the impact of transaction costs on mispricing. Ackert and Tian (2000) do not investigate the impact of idiosyncratic risk, since their sample does not facilitate a cross-sectional comparison. Their study still provides evidence on holding costs to the extent that SPDRs can be compared with closed-end funds. They find that SPDRs are subject to miniscule mispricing compared with closed-end funds. This finding is likely to be attributable to the low idiosyncratic risk associated with SPDR arbitrage, along with the fact that the SPDR redemption option will reduce the arbitrage holding period.

The importance of holding costs in these studies has an intuitive appeal, given the long-run nature of the abnormal returns. The longer the holding period of the arbitrage position, the more important holding costs are relative to transaction costs. In the case of closed-end funds, Pontiff (1995) shows that closed-end fund premia over net asset value (the mispricing measure) follow a slow mean-reverting process with a monthly autoregressive coefficient of 0.854. This implies that it takes, on average, 4.4 months for discounts to revert by half of the original discount, and 8.8 months for a $\frac{3}{4}$'s reduction. Regarding the predictability of book-to-market and accruals, Hirshleifer et al. (2004), show that both variables predict the cross-section of returns for at least 3 years. Regarding long-run SEO mispricing, Loughran and Ritter (1995), find post-SEO underperformance for holding periods of, at least, 5 years. Bernard and Thomas (1989) find that post-earnings announcement drift persists for about 6–9 months after the earnings release. No study has determined the holding period required to exploit the S&P 500 inclusion anomaly, although the findings of Shleifer (1986), imply that it is likely to be greater than 2 months.

Theoretically, idiosyncratic risk should be constructed relative to the universe of assets to which arbitrageurs have access. Because of this, return indices such as the S&P 500 are a natural choice as the independent variables that are used to compute idiosyncratic risk. As an empirical matter, as demonstrated by Mashruwala et al. (2006), the calculation of idiosyncratic risk does not seem to affect the results.

> *Myth 6*: Since arbitrage costs are related to mispricing, the costly arbitrage literature provides proof that markets are efficient.

Some people view costly arbitrage tests as showing that mispricing is "explained" by arbitrage costs. First, this view has already been discarded in the discussion of myth 1. Just because arbitrage costs are correlated with mispricing does not mean that mispricing is unprofitable; as long as holding costs exist, this correlation is consistent with the economic rationality of sophisticated traders, who trade until the marginal benefit of the position size is zero. Second, in a hypothetical environment with only transaction costs, arbitrage may still be unprofitable. Even in this extreme case, the finding of a relation between arbitrage costs and mispricing is still evidence of inefficiency. An efficient market is one where prices immediately reflect all relevant information. This definition says nothing about whether it is (or is not) profitable to trade against the inefficiency.

The most basic retort to myth 6 is that if the market is efficient, arbitrage costs should not be related to perceived mispricing. For example, a perceived market anomaly may be nothing more than the outcome of data-snooping. In this case, there is no reason to expect arbitrage costs to be correlated to the perceived mispricing—one would not find a relation between arbitrage costs and mispricing, and the test would be interpreted as showing that markets are efficient.

> *Myth 7*: Costly arbitrage tests are subject to Fama's (1976) joint hypothesis to the same extent as classic tests of market efficiency.

Costly arbitrage tests are tests of market efficiency. A classic test of market efficiency leads to Fama's (1976) joint hypothesis problem—rejection of the null of market efficiency may occur because the researcher is using the wrong asset pricing model or because the market is inefficient. As Pontiff and Schill (2004) show, costly arbitrage tests are useful because they test whether there is a specific relation between the apparent mispricing and costly arbitrage proxies. Unlike other behavioral tests, this relation is not ad hoc, since it is dictated by the behavior of economically rational traders. Even though mispricing may be measured with error, as long as the error is uncorrelated with arbitrage costs, the test is well specified. A costly arbitrage test is still subject to a joint hypothesis problem, but it is less likely to be an issue as it is for classic tests of inefficiency. In the case of relating abnormal returns to arbitrage costs, the joint hypothesis problem may be an issue to the extent that the finding may be a relation between expected returns and arbitrage costs. This is an unlikely possibility, especially for studies like Ali et al. (2003) and Mashruwala et al. (2006), since they show that the relation between arbitrage costs and returns is not monotonic—abnormal returns are increasing in arbitrage costs for apparently under-priced stocks, and decreasing in arbitrage costs for apparently overpriced stocks. All cross-sectional asset pricing models whether risk-, tax-, or transaction cost based, rely on monotonic relations between expected returns and the variables that drive expected returns.

6. Conclusion

Holding costs force arbitrageurs to take limited positions in mispriced securities. This enables mispricing to continue. Arbitrageurs are unable to hedge idiosyncratic risk, and thus they must trade off between the expected profit from a position and the idiosyncratic risk to which the position exposes them. The fact that idiosyncratic risk is an arbitrage cost is commonly misunderstood, and because of this, very few studies of market efficiency have examined the impact of idiosyncratic risk on mispricing. The empirical studies that have pursued this course share a common thread—idiosyncratic risk appears to be the single largest impediment to market efficiency.

Appendix A. Costly arbitrage and dividends

This appendix shows that for a given level of mispricing and expected mean-reversion in mispricing, arbitrageurs prefer positions that have higher dividend yields. This section only considers short-positions in which the fundamental value of the asset is risk free. The solution for long-positions, as well as positions with fundamental risk, yields similar results.

The fundamental value of a security at time t, is denoted by F_t. Without loss of generality, the fundamental value is riskless and has a zero expected return. Because of this, the level of the fundamental value is determined solely by last period's value and this period's dividend, D, thus

$$F_1 = F_0 - D. \tag{7}$$

The analysis starts by consider two mean-reverting mispricing processes in the absence of sophisticated traders. From these processes, we analyze the expected profit of an arbitrageur who does not affect prices, and how this expected profit is affected by dividend payments. Both processes that I consider assume that dividend payments do not affect mispricing.

A.1. Price level mean reversion

If the mispricing between price and fundamental value is related to the level of the difference, and if this mispricing follows a mean-reverting process with a zero long-run mean and a mean-reversion rate of θ, we can write,

$$P_1 - F_1 = \theta(P_0 - F_0) + \varepsilon_1, \tag{8}$$

where P_t is the price of the mispriced asset at time t and ε_t is the unexpected shock to mispricing at time t.

The arbitrageur will want to short sell the over-valued asset if $P_0 > F_0$. In this case, the expected profit from the short is the current value of stock, minus next period's expected value, minus the dividend that the arbitrageur must pay to the original owner of the share, minus the holding cost on the position. The holding cost is proportional to the level of next period's total arbitrage position, P_1, where the proportionality coefficient is h. Since the expected return on fundamental value is zero, the discount rate associated with this project is also zero. We can write expected profit of the position as

$$E(\pi) = (P_0 - E(P_1)) - D - hE(P_1). \tag{9}$$

Combining (7) and (8), taking the expectation and substituting it into (9), yields,

$$E(\pi) = P_0 - (1 + h)(F_0 - D + \theta(P_0 - F_0)) - D. \tag{10}$$

Differentiating the expected arbitrage profit with respect to the dividend payment yields,

$$\frac{\partial E(\pi)}{\partial D} = h. \tag{11}$$

Eq. (11) shows that holding all other variables constant, arbitrageurs earn higher profits from short-selling higher dividend yielding securities. Dividend payments decrease the expected future value of the position, thus expected holding costs are lower, and the position is more attractive.

A.2. Relative level mean reversion

Another potential mispricing process is mispricing that is related to the relative prices of the mispriced stock and its fundamental value. In this case, noise traders misevaluate the fundamental value by a certain percentage of value, as opposed to a dollar value (as implied by Eq. (8)). Assuming the same process for the fundamental value (Eq. (7)) and assuming that this process is mean-reverting with a zero long-run mean, we can write,

$$\frac{P_1 - F_1}{F_1} = \theta\left(\frac{P_0 - F_0}{F_0}\right) + \eta_1, \tag{12}$$

where η_1 is the unexpected shock to the mispricing process. Eqs. (12) and (7) imply that the next period price can be written as

$$P_1 = (F_0 - D)\left(1 + \theta\left(\frac{P_0 - F_0}{F_0}\right) + \varepsilon_1\right). \tag{13}$$

Using Eq. (9) definition of expected profit,

$$E(\pi) = P_0 - (1+h)(F_0 - D)\left(1 + \theta\left(\frac{P_0 - F_0}{F_0}\right)\right) - D. \tag{14}$$

Differentiating with respect to the dividend,

$$\frac{\partial E(\pi)}{\partial D} = (1+h)\left(1 + \theta\left(\frac{P_0 - F_0}{F_0}\right)\right) - 1. \tag{15}$$

Since the asset is over-priced, i.e., $P_0 > F_0$, higher dividend levels are associated with higher expected profits, even in the absence of holding costs ($h = 0$). Positive holding costs increase the incremental benefits of the dividend.

References

Ackert, L.F., Tian, Y.S., 2000. Arbitrage and valuation in the market for Standard and poor's depository receipts. Financial Management, 71–88.

Agnew, J., 2005. Inefficient choices in 401(k) plans: evidence from individual level data. Working paper. College of William and Mary.

Ali, A., Hwang, L.S., Trombley, M.A., 2003. Arbitrage risk and the book-to-market anomaly. Journal of Financial Economics 69, 355–373.

Asquith, P., Meulbroek, L., 1996. An empirical examination of short interest. Working paper, Harvard University.

Asquith, P., Mullins Jr., D.W., 1983. The impact of initiating dividend payments on shareholders' wealth. Journal of Business 56, 77–96.

Baker, M., Savasogul, S., 2002. Limited arbitrage in mergers and acquisitions. Journal of Financial Economics 64, 91–115.

Barber, B., Odean, T., 2000. Trading is hazardous to your wealth: the common stock investment performance of individual investors. Journal of Finance 55, 773–806.

Barberis, N., Thaler, R., 2003. A survey of behavioral finance. In: Harris, M., Stulz, R. (Eds.), The Handbook of the Economics of Finance. George Constantinides, North-Holland, Amsterdam.

Ben-David, I., Roulstone, D., 2005, Limited arbitrage and profitable trading: evidence from insider and firm transactions. Working paper, University of Chicago.

Bernard, V., Thomas, J., 1989. Post-earnings announcement drift: delayed price response or risk premium? Journal of Accounting Research 27, 1–36.

Bodie, Z., Kane, A., Marcus, A., 2004. Investments, 6th edn. McGraw-Hill Irwin, Boston.

Daniel, K., Hirshleifer, D., Subrahmanyam, A., 2001. Overconfidence, arbitrage, and equilibrium asset pricing. Journal of Finance 56, 921–965.

Dechow, P.M., Hutton, A.P., Meulbroek, L., Sloan, R.G., 2001. Short-sellers, fundamental analysis, and stock returns. Journal of Financial Economics 61, 77–106.

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

Duan, Y., Hu, G., McLean, R.D., 2005. Idiosyncratic risk and short-sellers: a test of the costly arbitrage hypotheses. Working paper, Boston College.

Fama, E., 1976. Foundations of Finance: Portfolio Decisions and Security Prices. Basic Books, New York.

Fama, E., French, K., 2005. Disagreement, tastes, and asset prices. Working paper, Dartmouth College.

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

Garman, M.B., Ohlson, J.A., 1981. Valuation of risky assets in arbitrage free economies with transaction costs. Journal of Financial Economics 9, 271–280.

Hirshleifer, D., 2001. Investor psychology and asset pricing. Journal of Finance 56, 1533–1597.

Hirshleifer, D., Hou, K., Teoh, S.H., Zhang, Y., 2004. Do investors overvalue firms with bloated balance sheets. Journal of Accounting and Economics 38, 297–331.

Jensen, M.C., 2005. Agency costs of overvalued equity. Financial Management 34, 5–9.

Lee, C.M.C., Shleifer, A., Thaler, R.H., 1991. Investor sentiment and the closed-end fund puzzle. Journal of Finance XLIV, 76–110.

Loughran, T., Ritter, J.R., 1995. The new issues puzzle. Journal of Finance 50, 23–51.

Markowitz, H., 1952. Portfolio selection. Journal of Finance 7, 77–91.

Mashruwala, C., Rajgopal, S., Shevlin, T., 2006. Why is the accrual anomaly not arbitraged away? The role of idiosyncratic risk and transaction costs. Journal of Accounting and Economics, in press, doi:10.1016/j.jacceco.2006.04.004.

Mendenhall, R., 2004. Arbitrage risk and post-earnings announcement drift. Journal of Business 77, 875–894.

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

Michaely, R., Thaler, R., Womack, K., 1995. Price reactions of dividend initiations and omissions: Overreaction or drift? Journal of Finance 50, 573–608.

Odean, T., 1998. Are investors reluctant to realize their losses? Journal of Finance 53, 1775–1798.

Petajisto, A., 2004. Why do demand curves for stocks slope down? Working paper, Yale School of Management.

Pontiff, J., 1995. Closed-end fund premia and returns: implications for financial market equilibrium. Journal of Financial Economics 37, 341–370.

Pontiff, J., 1996. Costly arbitrage: evidence from closed-end funds. Quarterly Journal of Economics 111, 1135–1151.

Pontiff, J., 1997. Excess volatility and closed-end funds. American Economic Review 87, 155–169.

Pontiff, J., Schill, M.J., 2004. Long-run seasoned equity offering returns: data snooping, model misspecification, or mispricing? Working paper, A Costly Arbitrage Approach, Boston College.

Pratt, J.W., Zeckhauser, R.J., 1987. Proper risk aversion. Econometrica 55, 143–154.

Samuelson, P.A., 1971. The fallacy of maximizing the geometric mean in long sequences of investing or gambling. Proceedings of the National Academy of Science, 2493–2496.

Schwert, G.W., 2004. Anomalies and market efficiency, Chapter 15. In: Constantinides, G., Harris, M., Stulz, R. (Eds.), Handbok of the Economics of Finance. North-Holland, pp. 937–972.

Shiller, R.J., 1984. Stock prices and social dynamics. In: Thaler, R.H. (Ed.), Advances in Behavioral Finance. Russell Sage Foundation, New York, pp. 167–207.

Shiller, R., 1989. Market Volatility. MIT Press, Cambridge, MA.

Shleifer, A., 1986. Do demand curves for stocks slope down? Journal of Finance 41, 579–590.

Shleifer, A., Vishny, R., 1997. The limits of arbitrage. The Journal of Finance 52, 35–55.

Stevens, G., 1998. On the inverse of covariance matrix in portfolio analysis. Journal of Finance 53, 1821–1827.

Thompson, R., 1978. The information content of discounts and premiums on closed-end fund shares. Journal of Financial Economics 6, 151–186.

Tuckman, B., Vila, J.-L., 1992. Arbitrage with holding costs: a utility based approach. The Journal of Finance 47, 1283–1302.

Wurgler, J., Zhuravskaya, E., 2002. Does arbitrage flatten demand curves for stocks? Journal of Business 75, 583–607.