The Limits of Arbitrage: Evidence from Exchange Traded Funds

Josh Cherry*

Department of Economics, University of California – Berkeley

In Partial Fulfillment of Honors in Economics

December 1, 2004

Abstract

Exchange Traded Funds (ETFs) consistently trade away from their net asset value. In violation of market efficiency, these discounts vary substantially over time and are found to be significant in the explanation of future returns. Returns to simple strategies which incorporate information in the variation of discounts outperform buy-and-hold strategies by an annualized 15%, net of transaction costs, but only expose the investor to about one fifth the risk. ETFs, on average, are found to be about 17% more volatile than their underlying assets; 70% of the excess volatility can be explained by proxies for transaction and holding costs which inhibit successful arbitrage. The findings in this paper are consistent with noise trader models of costly arbitrage and are inconsistent with hypotheses of financial market efficiency.

Keywords: Limits of Arbitrage, Efficient Markets Hypothesis, Exchange Traded Fund, Closed-end Fund Puzzle, Noise Trader Model.

JEL Classification: G10, G12, G14

*I would like to thank Botond Kőszegi for advising the preparation of this thesis, which would not exist without his invaluable comments, constructive criticism, and many helpful discussions. I would also like to thank Steve Goldman for direction and focus in the earliest stages. Lastly, I would like to thank Daniel Daneshrad for helpful comments and editing assistance.

1 Introduction

The efficiency of financial markets and the rationality of investors have long been the cornerstones of financial economics. As early as Friedman (1953), economists believed prices must reflect fundamentals because informed arbitrageurs could profitably eliminate any mispricings created by less informed investors. More recently, alternative theories of asset pricing in which arbitrage is not necessarily completed in the presence of sophisticated arbitrageurs have been put forward to explain empirical anomalies that are inconsistent with financial market efficiency.[1] For example, work by De Long, Shleifer, Summers and Waldman (1990), and Shleifer and Vishny (1997) have characterized conditions under which sophisticated traders are unable to profitably eliminate mispricings. Specifically, De Long, et. al. argue that mispricings will persist because noise traders can cause arbitrage to be prohibitively risky. Shleifer and Vishny specify a model in which arbitrageurs are constrained in their activities by agency problems.

This paper contends that the recent emergence of Exchange Traded Funds (ETFs) has provided a clear opportunity to test costly arbitrage theories. ETFs are unit investment trusts designed to replicate an index. The portfolios are highly observable since compositions of ETFs are published daily, and well diversified since ETFs follow indices. In similarity to closed-end funds, shares of ETFs are exchange traded. However, unlike closed-end funds, the supply of ETF shares is not perfectly inelastic; the trust is open ended in the sense that shares can be created and redeemed.[2]

For those reasons, ETFs are expected to be priced efficiently and to fit well in a classical model. I show that in actuality ETFs exhibit several properties that cannot be reconciled with the efficient markets hypothesis. Compared to closed-end funds, ETFs appear to be priced efficiently, that is to say the discounts are relatively small. ETFs, however, display discounts that are large considering their transparency and liquidity. Drawing from the closed-end fund literature, there are some structural characteristics that may explain the existence of relatively sizable discounts.

[1] Examples include short-run positive autocorrelation and longer-term negative autocorrelation, the closed-end fund puzzle, and the glamour-value anomaly.

[2] Creations and redemptions can only be executed in very large blocks called 'creation units'. Creation units vary in size from 25,000 to 600,000 shares. This characteristic is designed to prevent large deviations in the share price from the value of the trust.

The discounts may reflect for example the capitalization of future fund expenses (Brauer, 1988) or the relative liquidity of a fund to its assets.

Surprisingly, I find that ETF discounts exhibit more temporal variation than could be explained by the changes in the characteristics that are believed to cause them. Time varying discounts are shown to be a violation of market efficiency, yet I show ETFs exhibit significant variation in discounts over time. Furthermore, these variations in discounts are predictive of future returns. I construct simple trading strategies using information in the discounts and show that, net of fees, these strategies outperform the market substantially. This does not, however, mean that the arbitrage strategies are risky, as they require holding the ETFs on average only 17% of the time. Lastly, ETFs are demonstrated to be more volatile than their assets, an effect prohibited in a model based on investor rationality. The magnitude of both the abnormal returns and the excess volatility are related to the same proxies for transaction and holding costs.

One particularly well studied anomaly where costly arbitrage theories have been successful is the closed-end fund puzzle. Closed-end funds are investment companies that hold a portfolio of securities, and shares of this company are traded publicly. Surprisingly, these funds trade at a price that differs from value of the assets in the fund's portfolio. The difference in price is commonly referred to as the discount since closed-end funds typically trade at a discount.[3] Closed-end funds have been seen as the ideal vehicle for testing costly arbitrage theories because they generally hold observable, diversified portfolios. Pontiff (1996) analyzes a sample of closed-end funds in a costly arbitrage framework, and finds that costs associated with arbitrage explain about one quarter of the cross-sectional variation in discounts. The results therein are seen as a confirmation of costly arbitrage theories.

In this paper I will also argue there is reason to doubt that the results on closed-end funds confirm these new theories as well as ETFs. In several ways closed-end funds are not the best instrument for performing these tests. Closed-end funds are actively managed; consequently there is uncertainty about management ability. Furthermore, when management does change the composition of the portfolio, this is not immediately revealed to market participants. Closed-end funds

[3] A fund that trades at a premium is treated as a negative discount.

disclose holdings relatively infrequently, typically once a month or quarter. As investors are unsure of the portfolio composition, they become unable to price the funds correctly. These confounding effects, which do not apply to ETFs, cast suspicion on the previous confirmations of costly arbitrage theories.

The paper will proceed in the following way. Section 2 shows the excess volatility of the ETFs and estimates the relationship between transaction costs and excess volatility. Section 3 will discuss the time series properties of discounts and how they relate to returns. Section 4 will describe the arbitrage opportunity which results from these time series properties of discounts, and the relationship between transaction costs and abnormal returns. Section 5 will discuss the implications of these findings and conclude the paper.

1.1 Data

To estimate the time series properties of discounts and returns, and test for excess volatility, I use a daily sample of ETFs. There are altogether 83 iShares ETFs listed on the American Stock Exchange. Daily price and dividend data are obtained from *finance.yahoo.com*, and fund net asset value data are obtained from *www.ishares.com*. In this sample, Far-Asian funds update net asset value on the following business day. These data are susceptible to non-synchronous trading effects and are therefore excluded from the sample.[4] Funds incepted less than 100 trading days before February 3, 2004 are also excluded from the sample.[5] The sample ultimately contains 73 ETFs; both domestic and foreign stock funds as well as bond funds are included in the sample. All funds are considered from fund inception through February 3, 2004.

[4]These funds are MSCI Australia, MSCI Hong Kong, MSCI Japan, MSCI Malaysia, MSCI Singapore and MSCI Taiwan. The NAVs are updated at 10:00am PST for price changes on the previous business day.

[5]The funds incepted less than 100 business days prior to February 3, 2004 are iShares Lehman Aggregate Bond, iShares S&P TOPIX 150 Index, iShares Dow Jones Transportation Average, and iShares Lehman TIPS Bond Fund.

1.2 Returns and Discounts Definitions

To formalize returns and discounts define the investor return, the NAV return, and the discount of a fund as:

$$R_t^I = \frac{P_t + D_t - P_{t-1}}{P_{t-1}} \tag{1}$$

$$R_t^{NAV} = \frac{NAV_t + D_t - NAV_{t-1}}{NAV_{t-1}} \tag{2}$$

$$d_t = \ln\left(\frac{P_t}{NAV_t}\right) \tag{3}$$

where R_t^I is the investor return, R_t^{NAV} is the NAV return, P_t is the fund's price at t, D_t is the fund's dividend in t, NAV_t is the fund's NAV at t , d_t is the discount at t.

Funds that trade at a discount will therefore have $d_t < 0$, and a 'smaller' discount refers to a larger in absolute value discount or smaller in absolute value premium. The choice of the log discount ratio instead of discount levels does not change the results of this paper, but is important for two reasons. Firstly it simplifies calculations in testing for excess volatility in ETF markets. Secondly, it eases interpretation of coefficients in later regressions by capturing percentage changes in discounts.

2 Excess Volatility

DeLong, et. al., (1990) propose a model of limited arbitrage based on the existence of positive feedback traders who buy on price increases and sell on price decreases. Positive feedback trading can, for example, be generated by momentum strategies like trend-chasing or by the use of stop-loss orders.[6] DeLong, et. al., show that positive feedback can have a surprising effect on asset prices; it may not be rational for an arbitrageur to exert corrective price pressure towards fundamentals.[7]

[6]A stop-loss order is an instruction to sell a security at a price lower than the current price if the lower price is reached in trading.

[7]The consequences of the model are consistent with a number of stylized pricing irregularities. Asset markets typically exhibit short-term positive autocorrelations and long-term negative autocorrelations, which were thought to be anomalous but can be rationalized by the existence of positive feedback traders. Similarly, according to the model, asset prices should overreact to news, as they have been shown to do.

I contend that in the presence of positive feedback, price changes will be exaggerated relative to fundamentals (underlying securities). This means that positive feedback trading is expected to cause the variance of R^I to be higher than R^{NAV}. The reason to expect positive feedback trading in ETF markets is that they are frequently traded by individual (read: unsophisticated, uninformed) investors who are more likely to chase trends. This section will test the hypothesis that ETFs are more volatile than their NAV. While somewhat important in itself, the real motivation for investigating excess volatility is to make predictions about the time series relationship between returns and discounts. This section will demonstrate that ETFs are significantly more volatile than their NAV, and show that if a fund's return is more volatile than its NAV return, the fund's return will be strongly correlated with past discounts.

2.1 Testing for Excess Volatility

This section will demonstrate that if the volatility of R^I is larger than that of R^{NAV}, R^I will be strongly related to discounts over time. This suggests that the returns of funds with excess volatilty will be somewhat predictable by the discount. Pontiff (1997) develops with very modest assumptions a test for excess volatility that decomposes a fund's variance into the variance of discount changes and NAV variance.[8]

Proposition 1 *For a fund that pays zero dividends, $Var(R^I) > Var(R^{NAV})$ if, and only if,*

$$\frac{Cov\left(\Delta disc_{t,t-1}, R^{NAV}\right)}{Var\left(\Delta disc_{t,t-1}\right)} > -\frac{1}{2}.$$

If one were to regress R^{NAV} on $\Delta disc$, $\frac{Cov\left(\Delta disc_{t,t-1}, R^{NAV}\right)}{Var\left(\Delta disc_{t,t-1}\right)}$ has the natural interpretation as the coefficient of regression. Since $\Delta disc$ is the difference between R^I and R^{NAV}, if on average R^{NAV} decreases by less than half $\Delta disc$, R^I increases by more than half $\Delta disc$. This implies that R^I is more volatile than R^{NAV}. An important and testable prediction of this property is that whichever of R^I and R^{NAV} has more volatility will be more strongly correlated to $\Delta disc$ over time. I show

[8]Pontiff (1997) used the test on a sample of closed-end funds to show that they have 64% more volatility than their assets. The spirit of that paper motivates much of this section.

fund returns are more volatile than NAV returns to motivate the hypothesis that discounts will have predictive power for fund returns.

In this sample, 67 of 73 funds exhibit excess volatility relative to their NAV. Define the magnitude of excess volatility as $\frac{Var(R^I)}{Var(R^{NAV})}$. In this sample, the average (median) excess volatility is 17% (7%), (t-statistic of 39.68). To correct for possible skewness in the variance ratio, I also consider $ln\left(\frac{Var(R^I)}{Var(R^{NAV})}\right)$. The average (median) excess volatility is 15% (6.9%), (t-statistic of 6.75). To characterize the extent to which this can be surprising, the iShares MSCI Belgium Index Fund is 147% more volatile than its NAV.

2.2 Causes of Excess Volatility

Excess volatility persists due to limited arbitrage. Since there are several factors that limit the efforts of arbitrageurs, these factors should therefore affect in the same way the volatility of an ETF relative to its NAV. The inverse price should proxy for costs associated with bid-ask spreads, as cheaper securities tend to have larger relative spreads. A larger inverse price (relative spread) will increase excess volatility because trades are executed at relatively more disparate prices than is the NAV. Dividend yield is also important in explaining excess volatility. Arbitrageurs cannot fully invest short sale proceeds of dividend paying securities because the dividend accrues to whom the security was borrowed. Since traders are less willing to take short positions, there will then be less negative price pressure towards fundamentals, and hence more volatility. Interest rates are also expected to raise excess volatility, because the arbitrageur must bear the opportunity cost of his capital, and will thus be less willing to engage in arbitrage. Volume too should have an effect on excess volatility, but in the opposite direction of the other factors. Volume should increase liquidity and help mitigate the costs that prevent arbitrageurs from profitably causing convergence of price and NAV. To adjust for skewness in cross-sectional average volume, I take the log of average volume.

The asset class of a fund should also help explain the excess volatility. Bond funds should have lower excess volatility than equity funds because it is easier to price fixed income securities. International funds, however, are expected to have higher excess volatility because they typically have lower volumes and yields, and higher prices, and thus capture a lot of the variation in transaction costs.

I compute for each fund the inverse of the sample average price $pinv$, the average daily dividend yield div, the log of the sample average volume $lnVol$, and the average 13-week Treasury rate over each fund's sample period r_f. I also include dummy variables $intl$ and $fixed$ specifying international and bond funds, respectively.

Table 1
Cross Sectional Excess Volatility

Dependent Variable: $ln\left(\frac{Var(R^I)}{Var(R^{NAV})}\right)$				
Independent Variables	(1)	(2)	(3)	(4)
Constant	.242	.290	.266	.036
	(1.95)	(2.28)	(2.14)	(0.32)
Inverse Price, $pinv$	4.89	6.097	5.60	3.28
	(7.62)	(7.61)	(6.99)	(5.02)
Dividend Yield, div	4.05	—	4.34	2.82
	(2.19)		(2.36)	(1.59)
ln(Volume), $lnVol$	-.025	-.024	-.023	-.007
	(-2.30)	(-2.17)	(-2.11)	(-0.67)
13 Week Treasury, r_f	—	-.024	-.029	—
		(-1.18)	(-1.47)	
International, $intl$	—	—	—	.166
				(5.39)
Adjusted R^2	.6015	.5822	.6081	.7141
F-statistic	37.22	34.44	28.93	36.97

n = 73 *(Cross-Sectional regression of log excess volatility on Inverse Price, Dividend Yield, ln(volume), 13-Week Treasury Yield and International. t-statistics are in parentheses and are corrected for heteroscedasticity.)*

The excess volatility is not idiosyncratic. Table 1 shows these proxies for transaction and holding costs explain a large proportion of the variation in excess volatility exhibited cross-sectionally by ETFs. The slope coefficients in (1) are all of the predicted sign and statistically significant. Column (2) is the same regression as (1) with the exception that interest rates are used in lieu of dividend yield as a proxy for holding costs. All coefficients are of the predicted sign with the exception of interest rates, which are negative but not statistically significant (t-statistic of -1.18). Column

(3) uses both dividend yield and the interest rate to estimate the effect of holding costs on excess volatility, and the results are generally unchanged.

Column (4) includes asset class as an independent variable. The variable $fixed$ is not included in the reported regression because it is not significant, though the results are robust to alterative specifications which include it. It is interesting to note that in (4) the inclusion of *intl* has reduced the marginal effects of *pinv* and *div*. This occurs partly because international funds tend to have higher share prices and lower yields. Nevertheless, the specification in (4) appears to capture more variation in excess volatility. The coefficient on *intl* indicates that international funds have 18% more excess volatility than domestic funds.

The results reported in Table 1 suggest the inverse of price, dividend yield, and asset class are useful in explaining about 70% of the excess volatility, but that the effects of both interest rates and volume on excess volatility are likely spurious.

3 Estimating the Time Series Relations

This section estimates the relationship between time varying discounts and future returns for exchange traded funds. If a fund's discount has predictive power for future returns, it suggests funds prices are not informationally efficient. This would imply both investor irrationality and the existence of an arbitrage opportunity. This would certainly be difficult to reconcile with hypotheses of market efficiency.

Exchange traded funds are in many ways similar to closed-end funds. There are many explanations for the observed discounts of closed-end funds which, on first inspection, might be tractable for ETFs. These include transaction costs, the capitalization of future fees or uncertainty about management ability. These fund characteristics are relatively constant over time. If the deviations between price and NAV result from such fund characteristics, then for assets to be priced correctly any variation over time must be random. This section shows that not only are the variation of discounts over time is not random, they contain information about future fund returns. More formally, if $E(\Delta d_{t,t-1}) \neq 0$, then $E(R_t^I) \neq E(R_t^{NAV})$. This is obviously true because decomposing fund returns and taking the expectation of both sides yields:

$$E(R_t^I) = E(\Delta d_{t,t-1}) + E(R_t^{NAV}) \tag{4}$$

A simple argument following (4) should persuade the reader that a smaller than average discount should lead to a fund return that is higher than the NAV return, in expectation. This suggests discounts that vary over time violate market efficiency. Nevertheless, ETFs exhibit significant variation in discounts. The standard deviation of the discount for the average fund is 1.3%. An interesting example is the iShares MSCI Mexico Free Index that has an average daily discount of 4.4% and a daily discount standard deviation of 6.19%. Of the seventy three funds, twenty six exhibit a mean absolute daily discount larger than 50 basis points, and forty four exhibit differences larger than 25 basis points. These discounts are notably smaller in magnitude than that of closed-end funds, but the intraday variation in discount size is substantial nonetheless.

3.1 Model

I test the time series implications of the previous sections by estimating, for each fund,

$$R_t^I = \alpha + \beta disc_{t-1} + \epsilon \tag{5}$$

and

$$R_t^{NAV} = \gamma + \delta disc_{t-1} + \epsilon \tag{6}$$

where $disc_t = d_t - \hat{d}$.[9]

Result 1 *In equations (5) and (6) $\beta \leq 0$ and $\delta \geq 0$.*

If a different than average discount is a mispricing, the lagged discount should provide information about both future investor returns and NAV returns. Since it was shown in the previous

[9]The model is robust to alternative specifications. In particular, the residuals were also modelled as a fifth-order autoregressive processes. None of the autoregressive parameters were found to be significantly different from zero.

section that when $Var(R^I) > Var(R^{NAV})$, R^I will be more strongly correlated with discounts than will R^{NAV}. Therefore evidence for $\beta \leq 0$ is expected to be stronger than $\delta \geq 0$. This is intuitively appealing since it suggests the true fundamental value is somewhere between the ETF price and the NAV, but much closer to the NAV, and both converge towards it, on average. The NAV does not fully represent fundamental value since if traders choose to take quick positions in certain markets via ETFs, ETF prices can reflect fundamentals before the NAV does. When this happens the NAV will appear to move toward the ETF price.

3.2 Results of Estimation

Estimation confirms these expectations very strongly. For 71 of the 73 funds, β is of the predicted sign, and not significantly different from zero in the remaining two. Robust to heteroscedasticity, β is significant at the 1% level for 59 funds and at the 5% level for 67 funds. The mean (median) R^2 for (5) is 5% (3.4%), is greater than 5% for 24 funds, and is higher than 20% for two funds. This suggests the lag discount explains an economically significant part of the variation in daily returns. Similarly, the median β is -.68. The interpretation of the slope coefficient β is that a 1% higher discount lowers the next day expected market return by 68 basis points. In the most extreme, the smallest β is -1.25, indicating a very large effect on next day returns. Trading strategies based on β will be investigated in the next section.

For 49 of the 73 funds, δ is of the predicted sign, and not significantly different from zero when negative. The median δ is .06 indicating that in most funds, δ is not economically significant. Thus there is little evidence for NAV predictability, but the verification of $\delta \geq 0$ suggests weak evidence in favor of the hypothesis that NAV returns are positively related to discounts as predicted by the discussion of excess volatility.

4 Abnormal Returns Process

4.1 Mechanics of Arbitrage

Textbook arbitrage involves the simultaneous purchase and sale of identical assets at an advantageous price difference. For the present case that would involve, in a frictionless market, the purchase (sale) of an exchange traded fund and the simultaneous sale (purchase) of its underlying portfolio. Since exchange traded fund shares can be created and redeemed, rational speculators would continually cause the market price and NAV to converge.

There are, however, costs that will limit the ability of arbitrageurs to engage in classical arbitrage between ETFs and their representative portfolios. Firstly, costs associated with opening and closing arbitrage positions such as commissions, bid-ask spread, and market impact will limit arbitrage activity. Secondly, holding costs will be incurred every period by the arbitrageur. The arbitrageur must pay to borrow capital or bear the opportunity cost of employing his capital in arbitrage activities. There are also costs associated with the lost opportunity to fully invest short-sale proceeds. Casual inspection suggests classical arbitrage will not be profitable except for extreme mispricings.

It was shown previously in (5) that $\beta < 0$ and economically large. This suggests that traders may be able to earn abnormal returns by incorporating β from (5) in trading strategies. That is, a trader can base investment decisions on the predictive power $disc_{t-1}$ has for R_t^I. Since $\beta < 0$, a higher discount decreases the $E(R^I)$, and a lower one increases $E(R^I)$. A trader's strategy should purchase ETFs on very small discounts and sell them on very large ones.

4.2 Arbitrage Strategy

Define σ_d as the standard deviation of the daily discount and $\bar{d}$ as the mean d_t. For simplicity, let a discount d_t be considered *high* if $d_t \geq \hat{d} + a\sigma_d$, and considered *low* if $d_t \leq \hat{d} - b\sigma_d$, for some $0 \leq a \leq b$.[10] Here I relax the definition of arbitrage from its classic definition to the purchase

[10]The reason for the choice of $0 \leq a \leq b$ is not immediately obvious, but will become clear after the arbitrage strategy is defined more formally. The intuition is that it prevents the strategy from being too similar to a buy-and-

(sale) of a security with positive (negative) expected abnormal returns, without the simultaneous sale (purchase) of an identical or even highly correlated asset. Let a trader engage in arbitrage by purchasing an ETF with a *low* discount, and holding the security until the discount is *high*. The trader otherwise invests in the risk-free asset.[11]

4.3 Market Exposure

To formalize the returns to an arbitrage strategy, first define the market exposure function $\Phi(t)$:

$$\Phi(t) = \left\{ \begin{array}{lll} 1 & \text{if} & d_{t-1} \leq \bar{d} - b\sigma_d \\ 1 & \text{if} & d_{t-1} \leq \bar{d} + a\sigma_d \quad \text{and} \quad \Phi(t-1) = 1 \\ 0 & \text{otherwise} & \end{array} \right\} \tag{7}$$

The return to a strategy in t is therefore defined by:

$$\Psi(t) = (1 + R_t^I)\Phi(t) + (1 + r_f)(1 - \Phi(t)) \tag{8}$$

If the fund has existed for T trading days, and there are 250 trading days in a year, a buy-and-hold strategy will have returned an annualized rate of approximately $\left(\prod_{t \leq T}(1 + R_t^I)\right)^{250/T}$. The arbitrage strategy will have returned an annualized rate of approximately $\left(\prod_{t \leq T} \Psi(t)\right)^{250/T}$.

In this paper, I take $a = 1$ and $b = 2$.[12] The trader is invested in the ETF when $\Phi(t) = 1$. The trader will make all purchases and sales near 4 p.m. EST, which I will assume to have been executed at price equal to the closing price. That is to say the trader will invest in the ETF if

hold strategy (i.e. never triggering a 'sell' signal).

[11] The arbitrageur does not have to invest in the risk-free asset, but it is important that the alternate asset is uncorrelated with discounts. Chopra, et. al. (1993) suggest discounts on closed-end funds represent 'investor sentiment' and this measure of sentiment is correlated with the returns of stocks with low institutional ownership (i.e. small company stocks). If this were true of ETFs, one could not use the market, or a subset of it that is exposed to this investor sentiment effect, as the alternate asset.

[12] The results of this paper are fairly robust to the precise definition of *high* and *low* (i.e. the exact choice of a and b). The author is aware that more sophisticated definitions, and hence strategies, can be created and used to generate larger abnormal returns. The purpose of this paper is not to optimize over arbitrage strategies but to investigate the nature and cause of the violations of market efficiency.

its discount close to the end of the trading day is smaller than 2 standard deviations less than the mean, and will switch to being invested in the risk-free asset when the discount is higher than 1 standard deviation above the mean. The trader will switch from the risk-free asset back to the ETF when the discount is once again smaller than 2 standard deviations less than the mean.

4.4 Returns to Ψ

Firstly, $\overline{\Phi(t)} = .17$, that is the trader is exposed to the market only 17% of all trading days. Across funds, average (median) annualized abnormal return to this simple strategy gross of fees is 23% (7%). In the average (median) fund, the arbitrageur makes 3.5 (3.8) round trips per year, where a round trip is a purchase and subsequent sale of the ETF. Using 1% of assets as a conservative estimate for transaction costs, the arbitrageur expects to pay approximately $2 \cdot \#roundtrips \cdot 1\% \approx 8\%$ to execute the strategy. The net return will be $23\% - 8\% \approx 15\%$. Furthermore, this return is not risk-adjusted; the annualized abnormal excess returns are significantly higher.

There is substantial cross-sectional variation in the level of abnormal returns. Several factors are expected to have an effect on the level of abnormal returns earned by Ψ. These factors are related to the factors expected, and found, to explain excess volatility in a costly arbitrage setting. Firstly, bond and international funds are expected to have lower abnormal returns. Bond funds hold securities that are easier to price which leads to lower uncertainty in arbitrage activities, and ultimately more corrective price pressure. These activities should price away abnormal returns. International funds have on average higher volatilities and will exhibit less corrective pricing pressure and the distribution of discounts has fatter tails. In such a distribution, a discount considered *high* or *low* will occur relatively more often, and therefore such a discount will on average contain less information. More generally, the variation in discounts should negatively affect the magnitude of abnormal returns. Transaction and holding costs should have a positive effect on abnormal returns. In line with their effect on excess volatility, these costs prevent profitable corrective price pressure. Without the corrective price pressure the variation in the discounts will remain high and informative. Consequently, one can trade more profitably on the information contained therein.

In addition to the variables defined before, let $intl = 1$ if the ETF is international, $bond = 1$ if the ETF is fixed income, and let σ_{disc} be the standard deviation of the fund's daily discount. The dependent variable, annualied abnormal returns, is defined to be the excess returns (in percent) of the arbitrage strategy above a buy-and-hold strategy.

Table 2
Cross Sectional Abnormal Returns

Dependent Variable: Annualized Abnormal Return		
Independent Variables	(1)	(2)
Constant	3.54	3.62
	(3.95)	(4.28)
Inverse Price, *pinv*	3.84	—
	(0.59)	
Dividend Yield, *div*	-3.68	—
	(-0.23)	
ln(Volume), $lnVol$	-.279	-.286
	(-3.45)	(-3.81)
13 Week Treasury, r_f	-.006	—
	(-.04)	
International, *intl*	-.685	-.599
	(-2.77)	(-3.00)
Fixed Income, *bond*	-.243	—
	(-0.48)	
Discount Variation, σ_{disc}	-6.54	-6.04
	(-1.95)	(-2.36)
Adjusted R^2	.2047	.2372
F-statistic	3.65	8.46

n = 73 *(Cross-Sectional regression of annualized abnormal return on Inverse Price, Dividend Yield, ln(volume), 13-Week Treasury Yield, International, Fixed Income, and the Standard Deviation of Discounts. t-statistics are in parentheses and are corrected for heteroscedasticity.)*

The results summarized in Table 2 are all of the predicted signs. Inverse price, dividend yield and interest rates are found to be insignificant in explanation of abnormal returns. Surprisingly, abnormal returns to arbitrage in bond funds is not found to be significantly different from that of equity funds. More surprisingly, arbitrage in international funds is 60% less profitable even after controlling for the variation in discounts, despite the fact that the variation in discounts was thought to be the reason for lower arbitrage profits in international funds. This may instead be the result of the added foreign exchange risk of the underlying securities held by international ETFs.

4.5 Risk of Ψ

There is concern, however, that returns to Ψ are more risky than a buy and hold strategy. Heuristically, there is a temptation to say the risk must be lower since the same security is held, but for a subset of trading days. However, a more formal demonstration of this intuition is desirable. To empirically evaluate the risk of $\Psi(t)$, I estimate, for each fund, a single factor pricing model:

$$\Psi(t) - r_f = \alpha + \lambda \left(R^{NAV} - r_f\right) + \epsilon \tag{9}$$

A $\lambda < 1$ means that that Ψ is less volatile than the asset value of the fund. In this sample, the average λ is .19 (average t-statistic of 5.97) which suggests Ψ is about one fifth as risky as R^{NAV}. Since R^{NAV} is less volatile than R^I, Ψ must also be less volatile than R^I. Therefore, not only does Ψ outperform a buy-and-hold strategy, it is less risky. The fact that λ is so close to $\overline{\Phi(t)}$ is in no way surprising: it is a product of the proportionality of market exposure time to risk.

5 Discussion

5.1 Implications

The results of this paper have serious implications for financial market efficiency and investor rationality. If investor rationality characterized financial markets, arbitrage would be complete and ETFs would not be excessively volatile. Though there is a good indicator of fundamental value, the pricing of ETFs are still inconsistent with hypotheses of market efficiency. For securities like common stocks where fundamental value is much harder to determine, this paper suggests they too exhibit excess volatility and are likely to be inefficiently priced. The equity premium, widely believed to be too large, may not be. The equity premium may be able to be decomposed into a risk premium associated with fundamentals and an added premium for the excess volatility arising from investor irrationality. The attempt is left to future research.

While the results of this paper suggest investors are irrational, it says nothing about the systematic mistakes underlying these behaviors. Though speculative, one possible mechanism is that investors in these markets infer too much from small pieces of information. If, for example, trend-chasing ETF investors are trying to forecast price instead of fundamental value, observing price changes might be mistakenly over-informative. Again, the attempt to identify the precise mistakes in the decision process of investors is left to future research.

5.2 Conclusions

This paper characterizes a breakdown in investor rationality and under what conditions it is most likely to occur. ETFs are shown to be a better place than closed-end funds to test the costly arbitrage theories because they are more transparent, liquid, and are not subject to uncertainty about management ability. Unlike much of the literature, this paper focuses on the variation in the discount over time, as opposed to the size of the discount itself.

Exchange traded funds exhibit 17% excess volatility. Within a costly arbitrage framework, about 70% of the cross-sectional variation in excess volatility can be explained by the inverse price, dividend yield and asset class of a fund, all of which proxy for transaction costs. Excess volatility also implies that a fund's return will be strongly correlated with changes in discounts. Different than average discounts do, in fact, explain on average 5% of the variation in next-day returns. This predictive power can be exploited with simple arbitrage strategies to generate abnormal returns in excess of 15% per anum, net of transaction costs, and risk-adjusted returns that are significantly higher. The returns to such a strategy are lower cross-sectionally for funds with higher volumes and higher variation in discounts, and when the fund holds international securities. Since the extent of the pricing anomaly is strongly related to transaction costs, these results are consistent with theories of costly arbitrage.

Appendix: Proofs

Proof of Proposition 1. After Pontiff (1997), if $R_t^I = \Delta disc_{t,t-1} + R_t^{NAV}$, then the return variance can be decomposed as:

$$Var(R_t^I) = Var(\Delta disc_{t,t-1}) + Var(R_t^{NAV}) + 2Cov(\Delta disc_{t,t-1}, R_t^{NAV}), \tag{10}$$

$$Var(R_t^I) > Var(R_t^{NAV}) \Leftrightarrow Var(\Delta disc_{t,t-1}) > -2Cov(\Delta disc_{t,t-1}, R_t^{NAV}) \tag{11}$$

The result follows immediately. □

Proof of Result 1. Decomposing fund returns yields and taking the expectation of both sides:

$$E(R_t^I) = E(\Delta d_{t,t-1}) + E(R_t^{NAV}), \tag{12}$$

which implies

$$E(R_t^I) = E(d_t) - E(d_{t-1}) + E(R_t^{NAV}). \tag{13}$$

But $E(d_{t-1}) = d_{t-1}$ since it is known in $t-1$, and therefore

$$E(R_t^I) = E(d_t) - d_{t-1} + E(R_t^{NAV}). \tag{14}$$

In this formulation it is easy to see that all else equal R_t^I is decreasing in d_{t-1} and R_t^{NAV} is increasing in d_{t-1}. Therefore a regression of R_t^I on d_{t-1} is expected to have a negative slope coefficient, and in a regression of R_t^{NAV} on d_{t-1}, the slope coefficient should be positive. □

References

1. Berk, Jonathan B.; Richard Stanton, 2003, A Rational Model of the Closed-End Discount, Working Paper, available at:
http://faculty.haas.berkeley.edu/stanton/papers/pdf/closed.pdf

2. Bourdeaux, K. J., 1973, Discounts and Premiums on closed-end mutual funds: A study in Valuation, *Journal of Finance* 28, 515-522.

3. Brauer, Gregory A. 1988, Closed-end fund shares' abnormal returns and the information content of discounts and premiums, *Journal of Finance* 43, 113-128.

4. Chen, Nai-Fun; Raymond Kan; Merton H. Miller, 1993, Are the Discounts on Closed-End Funds a Sentiment Index?, *Journal of Finance* 48, 795-800.

5. Chopra, Navin; Charles M. C. Lee; Andrei Shleifer; Richard H. Thaler, 1993, Yes, Discounts on Closed-End Funds Are a Sentiment Index, *Journal of Finance* 48, 801-808.

6. De Long, J. Bradford; Andrei Shleifer, 1991, The Stock Market Bubble of 1929: Evidence from Closed-End Mutual Funds, *Journal of Economic History* 51, 675-700.

7. De Long, J. Bradford; Andrei Shleifer; Lawrence H. Summers; Robert J. Waldmann, 1987, Noise Trader Risk in Financial Markets, NBER Working Paper 2385.

8. De Long, J. Bradford; Andrei Shleifer; Lawrence H. Summers; Robert J. Waldmann, 1990, Positive Feedback Investment Strategies and Destabilizing Rational Speculation, *Journal of Finance* 45, 379-395.

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

10. Lee, Charles M. C.; Andrei Shleifer; Richard H. Thaler, 1991, Investor Sentiment and the Closed-End Fund Puzzle, *Journal of Finance* 46, 75-109.

11. Pontiff, Jeffrey 1996, Costly Arbitrage: Evidence from Closed-End Funds, *Quarterly Journal of Economics*, 111, 1135-1151.

12. Pontiff, Jeffrey 1997, Excess Volatility and Closed-End Funds, *The American Economic Review* 87, 155-169.

13. Poterba, James M., John B. Shoven, 2002, Exchange Traded Funds: A New Investment Option for Taxable Investors, MIT Department of Economics Working Paper 02-07.

14. Paul Samuelson, 1965, Proof that Properly Anticipated Prices Fluctuate Randomly, *Industrial Management Review* 6.

15. Shleifer, Andrei and Robert W. Vishny, 1997, The Limits of Arbitrage, *Journal of Finance* 52, 35-55.

16. Zweig, Martin E., 1973, An Investor Expectations Stock Price Predictive Model Using Closed-End Fund Premiums, *Journal of Finance* 28, 67-78.