Limited Arbitrage in the Secondary Market for Exchange-Traded Fund Shares

Doering, Philipp*

This Draft: December 7, 2017

Abstract

I study the profitability and determinants of relative mispricings between pairs of competing, nearly-identical Exchange-Traded Funds (ETFs) listed on US exchanges between 2007 and 2016. I find that prices sometimes diverge to an extent allowing to profitably trade on these deviations, with historical excess returns of up to 4 percent net of common fee estimates, suggesting considerable inefficiencies in the pricing of ETF shares. Price gaps are significantly larger among less liquid pairs, pairs with inactive primary markets, and on days with negative liquidity shocks. Though ETF pairs exhibit minimal convergence risk, arbitrage profits are positively related to holding costs as proxied by idiosyncratic risk. Altogether, common proxies for limits to arbitrage can explain up to 20 percent of the variation in arbitrage profits.

Keywords: law of one price, arbitrage, limits to arbitrage, market efficiency, Exchange-Traded-Funds, ETFs, pairs trading

* Ruhr-University Bochum, Department of Finance and Banking, Universitätsstrasse 150, 44801 Bochum, Germany; telephone: +49(0)234-32-21739, e-mail: philipp.doering@rub.de.

1 Introduction

With the ongoing shift from active to passive investing, Exchange-Traded Funds (ETFs) experience an increasing attention among both investors and academics. In 2015, assets under management surpassed hedge fund assets for the first time[1], and ETF shares are now accounting for about 30 percent of the overall US trading volume.[2] Besides providing low-cost access to diversification across all common asset classes, an often-cited key feature of ETFs is that they combine the benefits of both closed- and open-ended funds. Like closed-end funds, ETF shares can be traded intraday. Like open-ended funds, additional ETF shares can be created (or existing shares redeemed) through the "creation/redemption mechanism". The combination of these two characteristics provides a natural arbitrage channel: once the ETF share price diverges from the underlying net asset value (NAV), arbitrageurs buy the less expensive of both assets, convert it into the more expensive one and sell it, generating an (almost) immediate arbitrage profit (Ben-David, Franzoni and Moussawi 2016). This practice is also referred to as "primary market arbitrage", as it involves a change in the number of outstanding ETF shares.

Given this mechanism that by design intends to eliminate any mispricing within a short time, academics and practitioners long time paid little attention to potential ETF premiums and discounts. However, recent evidence suggests that assuming ETFs to always trade at their NAV may be a quite expensive mistake. Angel, Broms, and Gastineau (2016) argue that NAV deviations can be much greater than the bid-ask spread and thus, ETF transaction costs are often higher than investors realize. In fact, the aggregate of these hidden transaction costs is remarkable: in the US, investors pay approximately $40 billion each year for trading at premiums or discounts (Petajisto, 2017). On the one hand, there is evidence that these deviations can at least to some extent be linked to common limits to arbitrage (e.g., Madhavan and Sobczyk, 2016; Fulkerson, Jordan, and Riley, 2014), providing a rational explanation for the existence of premiums and discounts. On the other hand, even if deviations could be entirely explained by limits to arbitrage, these studies suggest that prices do not always reflect all available information, casting doubt on the efficiency of the ETF markets.

While focusing on price-NAV deviations as a measure of mispricing is certainly the most intuitive way to test the law of one price, another view is that competing ETFs, i.e. funds tracking the same benchmark, should sell at the same price. In other words, if the market for ETF shares is truly efficient, then it should neither be possible to profitably arbitrage ETFs against their underlying basket, nor against each other. In a perfect market without any impediments to arbitrage, there must be a linear combination in which the price spread between two competing funds is always zero, as otherwise, risk-free arbitrage profits would be possible. In real markets, of course, transaction and holding costs make arbitrage risky (e.g.,

[1] *"Stock Market Milestone: ETFs One-Up Hedge Funds As Investor Assets Hit $3 Trillion"*, Forbes Online, May 8, 2015.
[2] *„ETFs are eating the US stock market"*, Financial Times Online, January 1, 2017.

Pontiff, 2006), giving rise to temporary price deviations. Nevertheless, when ETFs are priced correctly, an arbitrage strategy in the secondary market for ETF shares should not allow to generate positive excess returns. This is the rationale underlying this paper. I select pairs of ETFs tracking the same benchmark, perform pairwise co-integration tests, and subsequently bet on price reversals between co-integrated funds. The high product homogeneity in the ETF market is somewhat unique and lends itself to view mispricings in relative terms. For example, as of April 4, 2017, NYSE Arca alone lists 45 ETFs tracking US technology stocks. Some funds are virtually duplicates: iShares, RBS, State Street and Vanguard all offer their own funds tracking the S&P MidCap 400 index. As illustrated in Figure 1, about 20 percent of all US-listed ETFs track indices that are also covered by at least one more fund. In value terms, approximately half of the total US assets under management are invested in funds that have at least one competitor tracking the same benchmark. This provides a fertile ground for arbitrage.

Viewing ETF pricing in relative terms is interesting for two reasons. First, as introduced by Petajisto (2017), considering an ETF's price distance to similar funds as a measure of mispricing rather than the distance to its NAV prevents the results being biased by NAV staleness. To provide an intuitive example, consider the SPDR S&P Russia ETF (ticker RBL). On April 17, 2014, RBL traded at a remarkable premium of roughly 350 basis points on its NAV. Was RBL actually mispriced? As the last NAV was recorded on Russian market close at 3:45 pm and the ETF closing price at 9:00 pm (UTC), the NAV lagged the ETF share price by approximately 5 hours. Within this timeframe, the Russian government agreed on a pact to defuse the Ukraine Crisis. This agreement is priced in the ETF share, but not in the last available NAV. Thus, the observed premium most likely reflected an informational gap, and focusing on the premium alone would have falsely suggested a mispricing. On the contrary, RBL was not mispriced in relative terms: the share prices of competing funds were all up by nearly the same amount.[3]

Second, evidence suggests that primary market arbitrage activities are rather scarce. Share creations and redemptions for a randomly selected ETF can only be observed on 6 to 13 percent of all trading days, and changes in ETF premiums or discounts are largely unrelated to prior share creations and redemptions (Fulkerson, Jordan, and Travis, 2017; Petajisto, 2017). Besides, there is evidence that primary market activity in a given ETF declines after new competitors enter the market (Box, Davis, and Fuller, 2016). These findings suggest that a substantial part of price correction happens in the secondary market alone, potentially even more when there are competing funds. Studying arbitrage opportunities between these funds might help understanding the role of secondary market arbitrage in enforcing efficient ETF prices.[4]

The contribution of this paper is twofold. First, I contribute to the literature on ETF pricing. While the idea of testing ETF prices in relation to competing funds is not new (see, in particular, Petajisto, 2017),

[3] Two competing ETFs are the iShares MSCI Russia and VanEck Vectors Russia ETF, tickers ERUS and RSX.
[4] Note that in the context of American Depositary Receipts (ADRs), where a conversion feature similar to the ETF creation/redemption mechanism exists, Alsayed and McGroarty (2012) find secondary market trading to be the major price-correcting mechanism to maintain stock-ADR price parity.

comprehensive evidence, covering a wide range of funds and properly studying the role of cross-sectional and time-varying limits to arbitrage simultaneously, is missing so far. For example, Marshall, Nguyen, and Visaltanachoti (2012) examine intraday mispricings between two S&P 500 ETFs. Fulkerson, Jordan, and Riley (2014) study whether similar bond ETFs can be arbitraged against each other. More recently, Petajisto (2017) used the cross-sectional average price of similar funds as a staleness-adjusted estimate for the NAV of a given ETF. He finds that the typical ETF trades in 100 bps range around this average price 95 percent of the time, implying that it is not unusual for competing funds to exhibit different prices. There are several important issues that have not been examined so far. First, are price deviations between competing funds really mispricings that can be profitably exploited net of transaction and other arbitrage costs? To which extent are these returns attributable to (other) cross-sectional and time-varying limits to arbitrage? How do potential mispricings emerge, i.e. what happens on the day prices diverge? How are they corrected subsequently? These are the questions I seek to answer. It is important to note that in contrast to most the aforementioned papers, I solely focus on pairs of near-perfect substitutes rather than "similar" funds.

Second, by addressing these issues, I also contribute to the more general literature on empirical asset pricing anomalies, where pairs of similar securities trade at different prices. For example, Schultz and Shive (2010) analyze price differentials between pairs of dual-class shares. Gagnon and Karolyi (2010) focus on price-parity among cross-listed shares. De Jong, Rosenthal, and Van Dijk (2009) study arbitrage returns in the context of dual-listed companies ("Siamese twins"). These papers have in common that though examining *close* substitutes, the paired-up securities are still exposed to some fundamental differences. Dual-class shares usually have different voting rights. Cross-listed shares in the US and their corresponding home-market shares as well as shares of dual-listed companies often trade in markets with different institutional features, such as disparately binding short-selling constraints, taxes, currency controls, or ownership limits (e.g., Gagnon and Karolyi, 2010; DeJong, Rosenthal, and Van Dijk, 2009; Froot and Dabora, 1999). The bottom line is that arbitrage profits can at least to some extent be attributed to fundamental risk, i.e. the risk that prices remain disconnected for an extended period of time. Two aspects make ETF pairs an interesting setting to study the profitability and limits to relative-value arbitrage. First, as already pointed out in Marshall, Nguyen, and Visaltanachoti (2012) and discussed in more detail later in this article, fundamental risk among ETF pairs should be minimized, if competing funds do not differ in their ability to track their benchmark index. At the same time, as I focus on pairs of assets trading in the same market, cross-market differences in institutional features should neither play a role. Second, in contrast to stocks, ETFs come with a two-tier liquidity structure, providing another dimension to examine how asset prices relate to liquidity (as will also be discussed in more detail in section 3).

The major results can be summarized as follows. First, though the magnitude of mispricings only averages to approximately 1 percent, they occur frequently enough for a profitable implementation of long-short arbitrage: trading on these deviations historically generated excess returns in the order of 2.5 to 4 percent p.a. net of fees. Pairs are typically equity ETFs, and the majority of funds considered in a pair portfolio is from the highest size and liquidity decile. Though 38 percent of all pairs employ different replication methodologies, fundamental risk is not a concern: arbitrage opportunities are typically triggered by a difference in premiums and discounts, while NAVs are quite close on the day of divergence, and pairs using explicitly different replication methodologies are not exposed to higher fundamental or other convergence risk than pairs with matching replication methods. Across all types of pairs, 82 percent of all price gaps converge within the defined trading periods, and prices are typically corrected within 4 days. Short sale constraints also play a negligible role in explaining why price deviations persist. However, I find a strong relation to cross-sectional differences in transaction costs, as arbitrage profits are substantially larger among more illiquid pairs. In particular, pairs with inactive primary markets, where a larger part of price correction is left to secondary market trading, tend to exhibit larger price gaps. Arbitrage profits are also related to idiosyncratic risk, tend to be larger on days with sudden drops in pair-level liquidity and on days with higher market-wide impediments to arbitrage. In combination, the limits to arbitrage proxies considered can explain up to 20 percent of the variation in arbitrage profits, providing a plausible explanation for price gaps.

The remainder of this article is structured as follows. The next section describes the ETF creation/redemption mechanism and provides a short review of the literature on ETF pricing. Section 3 briefly discusses the case for secondary market ETF arbitrage. Section 4 provides the sample and methodology employed. Results are presented in section 4 and section 5 concludes the paper.

2 ETF Pricing and the Creation/Redemption Mechanism

2.1 The Creation/Redemption Mechanism

In a frictionless market, an asset always trades at its fundamental value, as the concept of arbitrage implies that mispricings are corrected immediately. In real markets, however, arbitrage is limited to the extent that (i) both cognitive biases and constraints may impede information diffusion (e.g., Barberis and Thaler, 2003) and (ii) transaction and holding costs make arbitrage costly (e.g., Pontiff, 2006). Holding costs include the opportunity cost of capital, short-selling fees and idiosyncratic risk, with the latter often being considered as "the single largest cost faced by arbitrageurs" (Pontiff, 2006). The existence of holding costs implies that arbitrage in real markets is risky, as it makes the profitability of positions even in obviously mispriced assets contingent upon the time till convergence.

As mentioned in the introduction, the combination of intraday tradability and an open-ended structure underlying ETFs facilitates arbitrage activities. Specifically, primary market arbitrage is implemented as follows. Arbitrageurs monitor the price spread between the ETF share and the underlying basket. Once the spread gets too large, the arbitrageur buys the less expensive of both assets and short sells the more expensive one. At market close (09:00 P.M. UTC), the arbitrageur delivers the less expensive asset to the ETF sponsor in exchange for the more expensive one, covering the short sale and realizing an arbitrage profit at market close. Thus, the creation/redemption mechanism allows to exploit mispricings at minimum holding costs and, by design, aims to eliminate any observable price-NAV deviation in a short time. In a word, by creating or redeeming ETF shares in response to premiums, arbitrageurs can adjust the supply of ETF shares in a way that the fund trades close to the value of its underlying basket.[5]

In order to engage in the primary market arbitrage mechanism, i.e. to trade directly with the ETF's capital market desk, it is necessary to become an "Authorized Participant" (AP) by entering into an agreement with the fund sponsor first. An AP is typically a large institution, such as an investment bank or a broker-dealer. While most funds do not disclose their APs, estimates suggest that there are only a handful of institutions acting as APs worldwide.[6]

As shown by Petajisto (2017), share creations/redemptions are typically subjected to large minimum quantities between 50,000 to 100,000 shares, often requiring APs to accumulate their position over some days before submitting the creation/redemption order. Thus, in practice, they face uncertainty with respect to the timing of the simultaneous transaction in the underlying and the ETF as well as the costs associated with these trades (Petajisto, 2017). He finds that overall, the typical ETF only exhibits share creations/redemptions on between 6 and 13 percent of all trading days.

2.2 A Review of the Literature on ETF Premiums and Discounts

Academic literature on ETFs grew considerably in recent years. It can be broadly split into two different categories. First, there is controversy whether the increasing number of assets managed by ETFs may increase or reduce the efficiency of underlying security prices. Some studies suggest that the rise of ETFs enhances price discovery in the underlying markets (e.g., Madhavan and Sobczyk, 2016 and Glosten, Nallareddy, and Zou, 2016, to name a few). However, since ETFs played a major role in recent events of extreme market turbulence (such as the May 2010 Flash Crash), there are also concerns whether the creation/redemption mechanism may serve as a shock propagator. Indeed, there is evidence that due to their high liquidity and low trading costs, ETFs attract a clientele of short-term noise traders (Broman and Shum, 2016). Non-fundamental demand shocks caused by these noise traders may potentially be transmitted to underlying security prices through the arbitrage channel. Ben-David, Franzoni, and Moussawi (2017) provide evidence for this concern by showing that securities with higher ETF ownership exhibit higher non-fundamental volatility.

[5] An intuitive explanation on the creation/redemption mechanism is provided by the Investment Company Institute (ICI), see https://www.ici.org/viewpoints/view_12_etfbasics_creation.

[6] Financial Times Alphaville, "*Who exactly are authorised participants, anyway?*", https://ftalphaville.ft.com/2010/05/18/235011/who-exactly-are-authorised-participants-anyway/

The second strand of research is concerned with the pricing efficiency of ETFs themselves. Engle and Sarkar (2006) were among the first to study ETF premiums. Focusing on equity ETFs, they develop a statistical model to account for mismatches in timing between the last NAV and the last ETF share price. The overall finding is that once these mismatches are accounted for, premiums for ETFs tracking domestic benchmarks are generally small and short-lived. On the other hand, premiums for international equity ETFs are typically larger and often last for several days. Delcoure and Zhong (2007) support these results by showing that international ETFs trade at significant premiums even after controlling for differences in transaction costs. Ackert and Tian (2008) find that premiums of international equity ETFs are related to fund-level momentum, illiquidity, and size effects. Levy and Lieberman (2012) use intraday data to study mispricings of US-listed international equity ETFs. They show that the ETF share price follows the NAV during times of overlapping trading hours. However, in times the market of the underlying basket is closed, the ETF share price tends to follow the S&P 500. More recently, Hilliard (2014) and Angel, Broms, and Gastineau (2016) also observed higher and more persistent premiums among international ETFs, especially among funds tracking emerging markets.

The major part of the literature on ETF pricing is so far focused on share price-NAV deviations. However, there are some papers studying ETF prices relative to each other. First, Broman (2016) shows that premiums and discounts co-move across funds in similar investment styles. He argues that this is due to the correlated non-fundamental demand of noise traders, who are attracted by the relatively high liquidity of ETFs. Consistently, co-movements are stronger for funds with high commonality in demand shocks and attractive liquidity characteristics. Second, there are a number of papers specifically studying whether similar ETFs sometimes trade at different prices. Fulkerson, Jordan, and Riley (2014) study whether bond ETFs in similar investment categories can be arbitraged against each other. They find that a monthly rebalanced and equally weighted portfolio, buying the 10% of bond ETFs with the lowest premium and short selling the 10% with the highest premium, historically generated an alpha of approximately 11% per year before trading costs. More recently, Petajisto (2017) uses the cross-sectional average price of similar funds (though not necessarily funds tracking the same index) to estimate the “true”, staleness-adjusted NAV of an ETF on a given date. Though the focus of his research is to provide staleness-adjusted premium estimates, he implicitly shows that the typical ETF trades in a range between -50 and +50 bps around the average price of competing funds, implying that it is not unusual for similar funds to trade at different prices. Petajisto (2017) finally shows that a simple, daily rebalanced portfolio strategy, buying funds trading at a discount and short-selling similar funds trading at a premium, historically generated excess returns of up to 16% before trading costs.

My paper is probably closest to Marshall, Nguyen, and Visaltanachoti (2013), who use intraday data to study the microstructure of price deviations between two large and highly liquid S&P 500 ETFs. They find that there are only few and small-in-magnitude arbitrage opportunities, typically corrected within minutes. Mispricings are related to a fall in liquidity together with an increase in liquidity risk. Annualized, the profitability of exploiting these mispricings amounts to around 6 percent net of fees (but unadjusted for systematic risk).

In contrast to Marshall, Nguyen, and Visaltanachoti (2013), I use daily data, allowing me to cover a wide range of different funds, and thus to draw a more comprehensive picture on the efficiency of ETF prices and the role of both cross-sectional and time-varying limits to arbitrage. Are price deviations frequently and large enough to be traded profitably net of trading costs? To which extent can returns be linked to other limits to arbitrage, such as short-selling constraints, idiosyncratic risk, and illiquidity? For example, Petajisto (2017) does not address the contribution of leveraged and inverse ETFs to the returns of his strategy, though there is evidence that they are quite difficult to borrow (Avellanada and Dobi, 2013) and subjected to strict margin requirements.[7] Finally, what happens on the day price gaps emerge, and how are price deviations corrected subsequently? These are the questions I seek to answer.

3 Risks and Costs Involved with Secondary Market ETF Arbitrage

Secondary and primary market arbitrage are related to a different set of costs and risks. In particular, arbitraging ETFs against each other involves convergence risk. First, there is the risk that prices do not converge at all, because the funds may not be entirely identical. Compared to relative-value arbitrage in other settings (for example, dual-class shares), this fundamental risk (e.g., Mitchell, Pulvino, and Stafford, 2002) should be fairly low. As discussed in Marshall, Nguyen, and Visaltanachoti (2012), fundamental differences among ETFs tracking the same benchmark index are limited to only a handful a fund characteristics. For example, funds may use different methods to replicate their benchmark. While some funds physically buy all index constituents, others only hold a representative sample or employ a derivative-based approach. There may also be differences in the frequency in which dividends and other income received are reinvested or distributed to investors. Besides, some funds allow their securities to be lend to other market participants, while other funds do not participate in security lending activities. Luckily, whether a price deviation is fundamental or not can be measured in the context of ETF pairs by comparing NAV differences on the day of price divergence.

Second, even if prices certainly converge, it is ex ante unclear how long it will take (synchronization risk, see Abreu and Brunnermeier, 2002). In the meantime, noise traders may cause prices to diverge even further (see De Long et al., 1990), potentially forcing arbitrageurs to provide additional equity to their margin account or unwind the position. While these three risks play a negligible role in primary market arbitrage (as discussed in section 2.1), they are at least theoretically a concern when attempting

[7] For example, Interactive Brokers and Merrill Edge both multiply their margin requirements for common ETFs by the underlying leverage. Merrill Edge even prohibits trading ETFs that are leveraged three times or larger on margin.

to arbitrage ETFs against each other. Consequently, returns should be to some extent related to common proxies for holding costs, most importantly idiosyncratic risk (e.g., Pontiff, 2006).

Price deviations between competing funds should also be smaller among funds with higher liquidity, as liquidity is tied to transaction costs. Compared to other settings in which similar assets trade at different prices (e.g., dual-listed stocks), ETFs are somewhat special in that they have a two-tier liquidity structure. First, just like stocks, ETF shares can be traded in the secondary market, and the more actively the funds forming a pair are traded in the secondary market, the easier shares can be purchased or sold throughout the trading day. Second, even if "on-screen" liquidity is zero, ETF shares may be traded through the creation/redemption mechanism (see section 2.1). There is vast evidence that premiums and discounts (i.e. absolute mispricings) are related to underlying liquidity (e.g., Petajisto, 2017; Ackert and Tian, 2008; Engle and Sarkar, 2006), as more illiquid underlyings impede arbitrage through the primary market.

However, the impact of primary market liquidity on the persistence of *relative* price deviations between two competing funds is less straightforward and depends on the nature of mispricing. To the extent that both funds always have the same NAV and price gaps are solely due to diverging premiums or discounts, relative price deviations should ceteris paribus be more pronounced among fund pairs with illiquid underlyings. To provide some intuition, consider the extreme example of a fund with zero primary market activity. This could be either because the fund is always priced efficiently, or because primary market transactions are limited by underlying liquidity. In the latter case, price correction is entirely left to secondary market arbitrageurs. As secondary market arbitrage in contrast to primary market arbitrage involves convergence risk, relative price deviations should be larger among ETF pairs with illiquid underlyings in order to compensate for the additional arbitrage risk.

4 Data and Methodology

4.1 Data

I combine Morningstar Direct and Thomson Reuters Datastream to construct my sample. First, I use Morningstar Direct to obtain a list of all dead and alive US ETFs ever traded between 2007 and 2016. The choice of the sample period follows Petajisto (2017) and is thought as a compromise between the time period covered and the number of possible fund pairs. I limit my sample to funds listed on NYSE Arca, NASDAQ and BATS, as these are the major US trading places for ETFs, listing 1,973 of all 1,977 alive US ETFs (as of Dec 31, 2016). This initial sample covered a total of 2,531 dead and alive funds (dead: 558, active: 1,973). I then screened out a number of funds to obtain my final sample. First, I only retain primary shares to avoid pairs of different share classes. Second, as I utilize the funds' benchmark indices to form pairs, I remove all remaining funds that do not disclose a benchmark index in their prospectus. More precisely, I delete all funds which Morningstar classifies as "actively managed" or "enhanced index funds". I do however retain funds grouped as "strategic beta", which weigh constituents according to their factor exposures rather than market capitalization, as these funds always track a benchmark index. Overall, these filters reduced the sample size from an initial 2,531 to 2,262 funds. As of December 31, 2016, and across all share classes, the funds in my sample managed approximately $2.5 trillion, accounting for close to 99% of the total assets under management across all US ETFs. For the remaining funds, I again used Morningstar Direct to obtain other qualitative fund characteristics, such as style categories[8] and replication methods, as well as daily shares outstanding, total net assets, and NAVs. For funds that Morningstar classified as "leveraged" or "inverse" (317 funds in total), I hand-collected the corresponding leverage ratios from the fund prospectuses.

Second, I downloaded daily bid and ask prices, dividends and trading volumes using Datastream. For a total of 143 funds, Datastream has no coverage, further reducing the sample from 2,262 to 2,119 ETFs.[9] For the remaining funds, I compute daily bid-ask mid-quotes. Following the ETF literature (e.g., Engle and Sarkar, 2006; Broman, 2016), I use these throughout the paper, which also prevents my results being biased by the bid-ask bounce (e.g, Gatev, Goetzmann, and Rouwenhorst, 2006; Jegadeesh and Titman, 1995; Jegadeesh, 1990). In line with Petajisto (2017), I then define daily premiums and discounts as percentage difference of the daily mid-quotes from the corresponding NAV. I also follow the convention and subsequently use "premiums", even for negative observations (i.e. discounts).

To mitigate the effect of potentially erroneous quotes on my results, I apply a number of data filters closely following the related literature studying relative-value arbitrage in other contexts (e.g., Schulz and Shive, 2010; Marshall, Nguyen, and Visaltanachoti, 2012). For each fund, I discard all trading days where at least one of the following applies:

1. the bid quote, the ask quote, or both are missing,
2. the bid quote is equal to or greater than the ask quote,
3. the ask quote is exceeding the bid quote by more than 10 percent,
4. the ask or bid quote is below $5, as many brokers prohibit shorting "penny stocks".

Finally, I also remove all observations where premiums are larger than 20 percent in absolute terms, as these are likely to be erroneous (see Broman, 2016 and Petajisto, 2017). Trading is only allowed on the remaining days. Table 1 provides some sample characteristics.

[*Insert Table 1 here.*]

Panel A from Table 1 shows that at the end of 2016, the median ETF has $78 million assets under management. However, there is a large disparity: while the smallest fund has only $0.2 million in assets,

[8] Morningstar offers both Global and US categories. For my study, I use the US category classifications.

[9] Of the 143 removed funds, 53 were Exchange-Traded Notes. Besides, 95 of these 143 funds were active and 48 were dead as of December 31, 2016. Note that since my results base on long-short returns, survivorship bias is not a concern.

the largest fund (the SPDR S&P 500 ETF, ticker SPY) manages $225 billion. A similar pattern can be observed for trading activity and liquidity measures. For example, daily trading volumes range from zero to $24 billion for the most actively traded ETF. The most liquid ETF has an average spread of only 1 bp. On the other hand, there are also funds trading at spreads in the order of several percent.

Both the mean and median premium are close to zero, indicating that the typical ETF trades quite close to its NAV on an average day. However, premiums vary substantially over time: the median ETF exhibits a premium volatility of 46 bps, implying that 95 percent of the time, the typical ETF fluctuates in a range of -90 to +90 bps around its NAV. Thus, ETFs occasionally trade at an economically quite significant premium. These observations are quantitatively and qualitatively very similar to those by Petajisto (2017).

4.2 Methodology

A distinctive feature of ETFs is that similar funds can be easily identified by comparing benchmark indices. Thus, the basic idea to find ETF pairs is as follows. In the first step, I consider all possible pairs of funds tracking the same benchmark index as provided by Morningstar. While this should already result in a set of carefully pre-selected pairs, there are still two important implementation issues remaining.

First, the paired-up funds may employ a different leverage or even bet on different market directions. While this would not be an issue in case of a static ("set and forget") leverage, it is a major concern when considering to pair up leveraged or inverse ETFs. To provide some intuition, consider the example of (a) a two times leveraged position with static leverage and (b) an unleveraged position in the same asset, with both portfolios having an equity value of $100. In this case, portfolio (b) could always be replicated by holding a position worth $50 in portfolio (a) and $50 in cash. In the case of leveraged and inverse ETFs, however, this logic does not hold. The reason is that leveraged and inverse ETFs aim to deliver a multiple of the *daily* (sometimes monthly or quarterly) benchmark index return. Thus, these funds periodically reset their leverage, resulting in a volatility drag in the cumulative return of leveraged and inverse ETFs (e.g., Charupat and Miu, 2011; Jiang and Peterburgsky, 2017). As a result, prices of leveraged and inverse funds and their unleveraged counterparts certainly diverge over time, but for reasons other than mispricing. In other words, though tracking the same benchmark, they cannot be considered substitutes in price space. For this reason, I screen out all same-index pairs from the pre-selection, where the two funds either bet on different market directions or bet on the same direction, but employ a different leverage.

Second, the paired-up funds may charge different management fees. As these are subtracted from the fund's NAV pro rata on daily basis, prices will certainly disconnect over time in this case (see also Marshall, Nguyen, and Visaltanachoti, 2013). On the other hand, if both ETFs are indeed identical and do not charge different fees, potential prices deviations must be reverting to a constant mean.[10] Thus, I perform pairwise Engle-Granger co-integration tests (Engle and Granger, 1987). Only pairs that pass these tests are considered near-perfect substitutes and thus selected to be traded subsequently. The pairwise OLS regressions performed in the first step of these tests also allow to deal with the fact that similar ETFs are often divided into a different number of shares, implying that simply forming portfolios with equal share quantities in both legs may be inadequate.

More specifically, the matching algorithm is as explained below and inspired by the general literature on pairs trading (e.g., Jacobs and Weber, 2015; Gatev, Goetzmann, and Rouwenhorst, 2006). In accordance with this literature and to avoid a look-ahead bias, the algorithm is implemented rolling in two stages: pairs are formed based on 12 months of historical data (formation period) and traded in the subsequent 6-month period (trading period). The co-integration framework as outlined below follows Do and Faff (2016).

4.2.1 Formation Period

In each formation period, I first follow Petajisto (2017) and screen out the most illiquid funds, defined as having a daily average trading volume below $100,000 over the 12-month period. Of the remaining funds, I pre-select duplicate pairs as outlined above. I then perform co-integration tests for all pre-selected pairs. For this purpose, I use cum-dividend prices, i.e. cumulative total return indices with the initial index value set to the current ETF share price at the beginning of the formation period.[11] Based on these price series, I estimate the following model for each pre-selected pair k:

$$P_{k,1,t} = \alpha_k + \beta_k P_{k,2,t} + \epsilon_{k,t}, \tag{1}$$

where $P_{k,1,t}$ and $P_{k,2,t}$ are cum-dividend prices on day t for the two funds forming pair k. Augmented Dickey-Fuller tests are then applied to the residual, and all pre-selected pairs where the null hypothesis of non-cointegrated price series must be rejected at the 5%-level enter the final pair selection. For these pairs, the spread time series $\{P_{k,1,t} - \beta P_{k,2,t}\}$ is mean reverting. The estimated co-integration beta $\hat{\beta}_k$, as well as the historical price-spread mean $\hat{\mu}_{\epsilon,k}$ and standard deviation $\hat{\sigma}_{\epsilon,k}$, are recorded and serve as a trigger for opening and closing positions in the subsequent trading period (see also Rad, Low, and Faff, 2016).

[10] The presence of co-integrated price series is a necessary condition for two assets to be considered substitutes (Engle and Granger, 1987). While it is not a sufficient condition, the pre-selection based on qualitative criteria should lead to quite close substitutes.
[11] I use total return indices for testing on co-integration in order to account for differences in the distribution policy of funds. This inspired by the vast literature studying relative-value arbitrage in other contexts (e.g., Gatev, Goetzmann, and Rouwenhorst, 2006).

4.2.2 Trading Period

All pairs selected over the formation period are then eligible for trading in the subsequent 6-month trading period. During the trading period, the last recorded cum-dividend price from the formation period is updated based on current total returns. The normalized spread for a pair k on given day τ of the trading period, computed as

$$\frac{(P_{k,1,\tau}-\widehat{\beta}_k P_{k,2,\tau})-\widehat{\mu}_{\epsilon,k}}{\widehat{\sigma}_{\epsilon,k}}, \tag{2}$$

is monitored, and once it exceeds +2 or drops below -2, a long-short position is established. In general, which of the two funds has to be shorted depends on the sign of $\hat{\beta}_k$. However, as I screened out pairs of funds betting on different market directions, $\hat{\beta}_k$ can only be positive in my case. Thus, I buy one share of ETF 1 and sell short $\hat{\beta}_k$ shares of ETF 2, when the normalized spread drops below -2. If, on the other hand, the spread exceeds +2, I sell short one share of ETF 1 and buy $\hat{\beta}_k$ shares of ETF 2.[12] I close out the position when (i) the normalized spread returns to zero, (ii) a fund is delisted, or (iii) at the latest by the end of the current six-month trading period. When a pair completes a whole roundtrip within the trading period, it is eligible for another trade and subjected to the same methodology again. For the purpose of conservatism, it is assumed that a pair earns zero interest if it does not actually trade, i.e. capital not allocated to a pair is *not* invested at the risk-free rate.

It should be noted that in some cases, a price spread exceeding two historical standard deviations may only amount to a few basis points and thus be too small to cover trading costs. However, in order to avoid data snooping, I chose to follow simple instead of optimal trading rules. Whether price deviations are large enough to cover trading costs or not, I henceforth often refer to these observations as "mispricings", though the spread volatility may sometimes be too small to profitably trade on these deviations.

4.2.3 Return Calculation

As ETF pairs form long-short portfolios, computing returns is a non-trivial issue for two reasons. First, depending on the magnitude of $\hat{\beta}_k$, co-integrated pairs are not necessarily dollar-neutral. In a frictionless market, dollar-neutral portfolios are self-financing, i.e. the long position could be financed by the proceeds of the short sale. Second, even if pairs were dollar-neutral, real markets require arbitrageurs to post collateral for both long and short positions. Thus, I follow the vast literature concerned with arbitrage in other settings (e.g., Marshall et al., 2012; Schultz and Shive, 2010; De Jong et al., 2009; Mitchell, Pulvino, and Stafford, 2002) and calculate the return of a pair based on the capital that is required to

[12] For the sake of completeness: for pairs with $\hat{\beta}_k < 0$ (i.e. pairs that combine an inverse with a long ETF), one would buy both one share of ETF 1 and $\hat{\beta}_k$ shares of ETF 2, when the spread drops below -2. If the spread exceeds +2, one would have to short both one share of ETF 1 and $\hat{\beta}_k$ shares of ETF 2 (see Rad, Lo, and Faff, 2016).

bring up the position. Specifically, I assume that arbitrageurs have to meet Regulation T (Reg T) margin requirements. According to Reg T, investors are required to bring up 50 percent of the long and 50 percent of the short market value as initial margin.[13] Additionally, I assume a maintenance margin requirement of 30 percent for both the long and short position. For the sake of simplicity, I liquidate the position in a pair once the equity in either the short or long position drops below 30 percent of the position value.

Pair returns are then obtained by dividing the sum of payoffs from the long- and short-positions by the required equity. For the largest part of my analysis, I use net-of-fee returns, i.e. the payoffs considered in the nominator are net of spreads, commissions, short rebates and any interest paid on margin borrowing. Specifically, positions are marked-to-market daily by dividing the daily net-of-fee payoffs by previous day's equity. As I use mid-quotes to compute prices at both position entry and position close, bid-ask spreads are already accounted for. Commissions have been fairly low in recent years and are thus typically ignored in the related literature (e.g., Marshall, Nguyen, and Visaltanachoti, 2013). For example, investors with direct access to NYSE Arca are charged between 0.1 and 0.3 cents per traded share.[14] However, not all institutional investors are provided a direct access. Thus, I follow the more general estimates provided by the Investment Technology Group (ITG). According to ITG,[15] commissions average to approximately 5 bps for my sample period. Thus, as a whole roundtrip involves four transactions, commissions amount to 20 bps in total. Finally, as short rebate data are difficult to obtain for ETFs, I rely on the estimates provided by Stratmann and Welborn (2013), who report an average rebate rate of -1.13 percent per year for US ETFs. In other words, on average, ETF short sellers pay 1.13 percent per year to the lender, reflecting that ETFs are typically hard to borrow. As a proxy for the margin borrow interest on the long position, I use the Fed Open Rate on a daily basis plus 50bps (see, e.g., De Jong, Rosenthal, and Van Dijk, 2009). Though representative, these estimates may of course be inappropriately low or high for some ETF pairs. I will thus address the sensitivity of my results towards short selling constraints and liquidity in section 5.

Daily *pair* returns are then used to compute daily returns for the *portfolio* of all pairs. For this purpose, I assume that all pairs have the same weight at the beginning of the trading period. However, weights may change over time, since I assume that proceeds from previous trades within the same trading cycle are reinvested. I compute two different portfolio return measures: return on committed capital (ROCC) and return on employed capital (ROEC). ROCC adjusts the pair payoffs by the number of pairs that were selected for trading, while ROEC adjusts by the number of pairs that actually traded. To ease

[13] In some cases, the arbitrageur may decide to provide more capital in order to reduce leverage.
[14] NYSE Arca fees and charges can be found at https://www.nyse.com/publicdocs/nyse/markets/nyse-arca/NYSE_Arca_Marketplace_Fees.pdf.
[15] The preliminary version of the 2017 Global Cost Review is available at https://www.itg.com/assets/ITG_Global-Cost-Review-2017Q2-Prelim-BrokerCostUpdated.pdf.

interpretation, I follow the convention of the pairs trading literature (e.g., Gatev, Goetzmann, and Rouwenhorst, 2006) and compound daily portfolio returns to monthly returns before reporting.

As in the momentum and pairs trading literature, the above trading cycle is implemented in a manner that a six-month trading period begins every month (except in the first 12 months of the sample, which solely serve as an initial formation period for the first trading period). The result is a set of monthly return series for overlapping six-month trading periods, giving rise to six different return observations for each month. The actual monthly return on the pair portfolio is computed as the average across these six returns. While portfolio returns most appropriately capture the profits earned by an average arbitrageur trading on relative ETF mispricings, the impact of cross-sectional and time-varying limits to arbitrage can be more precisely studied based on single trades. The reason is that portfolios can be thin in some trading cycles, and in results unreported for brevity I found that portfolio size varies strongly over time. Thus, following the pairs trading literature (Engelberg, Gao, and Jagannathan, 2009; Jacobs and Weber, 2015), I report "position returns" in a large part of my analysis instead, which are simply the returns of single trades (either gross or net of fees).

5 Results

5.1 Descriptive Pairs Statistics and Co-Integration Parameters

Table 2 reports the pair frequency, characteristics of paired-up ETFs, and pairwise co-integration statistics. As can be seen from Panel A, from a total of 4,476 eligible pairs, about 3,677 were found to be co-integrating. On average, 39 out of 32 possible pairs were selected for trading each month. Hence, about 82 percent and thus the vast majority of pre-selected pair combinations actually co-integrate.

[insert Table 2 here]

Panel B shows descriptive statistics for co-integrated pairs. The key insights can be summarized as follows. First, the average ETF matched into a pair has about $7 billion in net assets and is thus from the highest size quartile (see Table 1). The median amounts to a substantially smaller $636 million, which is still in the top quartile. Second, compared to all funds, the typical pair constituent is from the highest liquidity quartile, with a median bid-ask spread of merely 5 bps and a daily trading volume of $4 million. Besides, pair-funds exhibit close-to-zero premiums and a below-median premium volatility of 33 bps on average. Premiums of paired-up funds typically exhibit a correlation of 0.4 and are significantly correlated in 65 percent of all cases. These results suggest a co-movement in premiums among competing funds as in Broman (2016).

Panel C provides insights into the estimated coefficients of the co-integration regression presented in equation (1), performed during the formation periods. As can be seen from the co-integrating beta, the number of shares to be held in both ETFs to obtain the mean-reverting pair portfolio is quite balanced. For the typical pair, the ratio of shares to be held in both ETFs is approximately 1-to-1 (as measured by the median). The average is somewhat higher and suggests a typical ratio of 1-to-1.6. The mean price-spread under the co-integrating relationship is close to $1 for the average pair. Thus, the price difference between the two funds in a typical pair is small in dollar terms. The residual volatility is close to $0.50 for the average pair, implying that the price spread fluctuates within a 100-percent range around the $1 mean 95 percent of the time. Besides, the price spread crosses its time-series average on 24% of all trading days. Finally, the last row shows that of all pairs selected during the formation period, 75 percent remain co-integrated over the subsequent trading period, implying that a long-short strategy as outlined in section 4.2.2 should – at least gross of fees – generate positive returns.

5.2 Risk and Return

Table 3 summarizes the portfolio returns for the arbitrage strategy outlined in section 4.2. Panel A shows that regardless of the portfolio return measure used and whether fees are accounted for or not, the strategy exhibits both statistically and economically significant positive returns. When adjusting using employed capital, net-of-fee returns average to 36 bps per month, which is an annualized 4.4 percent. Using the more conservative approach to compute portfolio returns, i.e. adjusting with committed capital, leads to an average net-of-fee return of 27 bps per month (annualized: 3.3 percent), which is still significant in both economic and statistical terms. Unsurprisingly, gross-of-fee portfolio returns are 8-12 basis points larger. In all cases, returns remain significantly positive after subtracting the risk-free rate.

[insert Table 3 here]

The overall market return (gross of fees) averages to 67 bps per month. Compared to the most conservative profitability measure for the ETF arbitrage strategy, the net-of-fee return on committed capital, the average market return is twice as large. However, the strategy exhibits a substantially lower risk, regardless of the risk measure used. The realized market return volatility is 458 bps per month, whereas the strategy's return volatility amounts to a mere 70-100 bps. This disparity persists when adjusting with downside risk measures. For example, the strategy experienced negative net-of-fee returns in only 22 percent of all months, whereas a buy-and-hold market investor suffered losses 40 percent of the time. Consequently, as reported in Panel B, the strategy outperformed the market, regardless of the performance measure used.

[insert Table 4 here]

Table 4 shows that only a small portion of the excess returns reported in Table 3 can be attributed to common equity risk factors. As a market-neutral strategy, arbitrage profits are not significantly exposed to the equity risk premium. Exposures to other factors are mostly insignificant, with two exceptions: returns are weakly related to the conservative-minus-aggressive factor and load negatively and significantly on the momentum factor. Nevertheless, regardless of the return measure and factor model used, net-of-fee alphas are statistically and economically significant, ranging from 21 bps to 33 bps per month (annualized: between 2.5 and 4 percent). It should be noted that the magnitude of returns is similar to the results obtained by Marshall, Nguyen, and Visaltanachoti (2012) for intraday arbitrage between two S&P 500 ETFs, who report returns in the order of 6 percent per year (unadjusted for systematic risk). On the other hand, these returns are considerably smaller than those found in previous papers using daily data without accounting for trading costs. For example, Petajisto (2017) finds excess returns in the order of 16 percent per year. Fulkerson, Jordan, and Riley (2014) focus on bond ETFs and report an historical alpha of approximately 11 percent per year. Nevertheless, the fact that we observe systematic excess returns even net of fee estimates indicates that there are noticeable inefficiencies in the pricing of ETFs.

Using equity factors to adjust for systematic risk would be inappropriate if the majority of selected pairs were non-equity ETFs. Thus, Table 5 provides several information on the composition of pair portfolios selected for trading, averaged across all trading cycles. From Panel A it can be seen that about two-thirds of all funds selected in a pair portfolio are from the largest three size deciles, as measured by the total net assets at the end of the preceding formation periods. 85 percent of portfolio constituents record above-median net assets. Similar proportions can be observed in terms of liquidity: the majority of funds considered in a typical portfolio trades at low bid-ask spreads relative to all ETFs investable at that time. Thus, on average, portfolios primarily consist of relatively large and liquid funds. However, at the same time, about three-quarter of all pairs consist of ETFs from different size and liquidity deciles, reflecting that for many indices, there is one market-leading fund and a number of smaller competing products. On average, paired-up funds differ by 2.7 size and 2.5 liquidity deciles.

[insert Table 5 here]

Panel B reports the weight of pairs according to their underlying asset class and index type. 80 percent and thus the vast majority of all pairs selected for trading consist of funds tracking stock indices. The remaining 20 percent are divided equally in commodity and fixed income pairs. Close to 40 percent of all pairs are comprised of funds tracking factor-weighted instead of traditional cap-weighted indices, reflecting the substantial growth in "smart beta" products in recent years. Panel C shows that paired-up funds are typically competing, i.e. ETFs issued by different sponsors, and the majority of pairs (62 percent) employs the same methodology to replicate their underlying indices. On the other hand, the latter implies that 38 percent of all pairs certainly hold different security baskets to track their benchmark indices. In this case, there is the risk that both funds sometimes differ in their tracking ability, implying a fundamental risk of permanent price disconnects.

[insert Table 6 here]

However, in line with matching replication methodologies for the majority of pairs, Table 6 shows that price deviations are rather driven by diverging premiums than NAV discrepancies. While the total price deviation between the short and the long leg on the day positions are opened averages to 1.2 percent, only 16 bps can be attributed to NAV differences. Premiums differ by an average of 95 bps on position opening. It should be noted that price discrepancies are right-skewed, as the median difference is substantially smaller in all cases. Nevertheless, the observation holds that price gaps are typically driven by diverging premiums rather than diverging NAVs. Besides, prices subsequently converge by a decay of differences in premiums, whereas NAV distances remain almost constant, and changes in average NAV differences are both insignificant in economic and statistical terms.

[insert Table 7 here]

Table 7 provides insights into the frequency and profitability these price disconnects. From a total number of 5,308 price deviations, 82 percent converged within the six-month trading periods. The median time-till-convergence of 4 days suggests that price gaps are corrected quickly, compared to other settings where similar assets trade at different prices. For example, prices of dual-listed companies typically remain disconnected for 22 days, and in rare cases the time till convergence amounts to several years (De Jong, Rosenthal, and Van Dijk, 2009). The average convergence time is 10 days and thus considerably higher than the median, implying that in some ETF pairs, mispricings persist for an extended period of time. On average, about 60 percent of all pairs trade during the six-month trading period, and there are typically 1.4 arbitrage opportunities within an average pair. If a pair opens, it experiences two roundtrip trades on average. The average gross-of-fee return of single positions averages to 1.26 percent, while net-of-fee returns are around 30 bps smaller on average. Losses occur in only about 25 percent of all trades and are typically small. However, as can be seen from the first percentile, prices can in rare cases diverge further, involving a loss of 1.5 percent or larger. There is also a large disparity among profitable trades: while the median return is 60 bps net-of-fees, 25 percent of all trades generate a net return exceeding 1.8 percent.

5.3 The Role of Cross-Sectional Limits to Arbitrage

5.3.1 Fundamental Risk

Considering the high convergence probability and quite limited losses, combined with the observation that NAV deviations are typically small on days where relative mispricings occur, it can be inferred that fundamental risk is negligible among pairs of similar ETFs. To provide further insights, I split all trades into those executed within pairs using the same and pairs using explicitly different replication methodologies. If fundamental risk were a major concern, non-converging, loss-making trades should occur more frequently and be more pronounced among pairs employing different replication methodologies. However, the results reported in Table 8 do not support this notion. While pairs of funds employing different replication methodologies indeed exhibit a significantly lower return, this difference is rather attributable to a lower profitability of converging trades than more frequently striking arbitrage risks, implying that mispricings are typically smaller among these pairs. The average loss of unconverged trades is significantly larger among same-replication pairs, and mispricings more likely converge among pairs using different replication methodologies. Besides, the standard deviation of position returns is significantly smaller.

[insert Table 8 here]

Overall, these results suggest an overall lower convergence risk for fund pairs with different replication methods, and thus point in the opposite direction of what would have been expected if these pairs were exposed to considerable fundamental risk. I infer from these results that fundamental risk cannot be a major concern discouraging arbitrageurs from eliminating mispricings.

5.3.2 Short Sale Constraints

While the baseline analysis already assumed a short rebate rate of -1.13 percent per year, considering that ETFs are typically hard to borrow (see Stratmann and Welborn, 2013), there is anecdotal evidence that short selling may be even more difficult or sometimes impossible for some types of funds. In particular, Avellanada and Dobi (2013) report that the costs for borrowing leveraged and inverse ETFs are rather ranging from 250 to 850 bps. Besides, in 2009, FINRA implemented stricter margin requirements for leveraged and inverse ETFs.[16] Some brokers even entirely exclude a subset of these funds from margin trading.[17] Similar concerns hold for margined positions in Exchange-Traded Notes (ETNs), which in contrast to ETFs are debt instruments and do not involve a claim in an underlying pool of assets.[18]

[insert Table 9 here]

If short sale constraints can explain why price gaps persist, deviations should be driven by an overpricing in hard-to-borrow ETFs, and returns should be primarily stem from the short leg of the position (e.g., Gatev, 2006). Panel A of Table 9 shows that this is indeed the case for price deviations in leveraged or inverse funds. The net-of-fee return, averaged across all trades in leveraged or inverse pairs (subsequently "leveraged" for brevity), amounts to a statistically significant 134 bps. When looking at the long and short leg separately, it can be seen that only the short leg experienced a significant 244 bps return on average, while the long leg on average suffers a (though insignificant) loss. In contrast, trades executed within pairs of unleveraged funds tracking the same indices covered by at least one leveraged fund pair exhibit a significantly different return profile. For these trades, returns are primarily driven by the long leg.

[16] See Regulatory Notice 09-53, effective as of December 1, 2009. As a consequence, margin requirements for leveraged and inverse funds increased by a factor commensurate with their leverage.
[17] For example, Merrill Edge prohibits trading ETFs that are leveraged three times or larger on margin.
[18] For ETNs, the issuer typically alone has the right to decide on the number of outstanding shares.

For ETNs, however, Panel B suggests that short sale constraints are not a major impediment to arbitrage. For both, pairs involving at least one ETN, and pairs exclusively consisting of ETFs, returns primarily stem from the long leg. Short position returns are even significantly negative in either case. Panel C shows that the results do not significantly change when excluding leveraged pairs and pairs involving ETNs, which is unsurprising given that trades in these pairs represent only about 2 percent of all identified mispricings. Besides, looking at the long and short leg of the remaining trades separately reveals that arbitrage profits can on average be attributed to the long position, while short position returns are even negative on average. Thus, pair returns are typically driven by an underpricing in one of the two ETFs, and short selling constraints cannot explain why relative mispricings between similar funds persist.

5.3.3 Cross-Sectional Differences in Holding and Transaction Costs

In order to test whether price discrepancies are related to arbitrage costs, I construct a number of holding and transaction cost proxies. First, while fundamental risk is minimized, trading on relative ETF price deviations does still involve some convergence risk. As discussed in section 5.2, 18 percent of all deviations do not converge within the six-month trading periods, and even for pairs that do converge within this time frame, it is ex ante unclear how many days it will take for prices to be corrected. Thus, though limited, arbitrageurs do face holding costs, and returns should be related to common holding cost proxies. The most important and most frequently used holding cost proxy in the literature is idiosyncratic risk (e.g., Pontiff, 2006), reflecting the risk of the arbitrage position that is unrelated to systematic factors and any other available hedge portfolio. It is far from clear how to precisely measure the idiosyncratic risk of an arbitrage position (see Gagnon and Karolyi, 2010). The approach I use in this paper closely follows Jacobs and Weber (2015), who use pair-level averages of residual standard deviations from factor regressions as idiosyncratic risk proxy associated with a long-short position in related stocks. However, in the case of ETFs it should be noted that a fundamental part of fund returns could theoretically be hedged holding the underlying index. Thus, I only consider the differential between ETF and NAV returns. The resulting excess return series are regressed on the corresponding Fama and French (1992) factors. I then compute for each fund the residual standard deviation from these regressions. Finally, equally weighted pair-averages across individual residual standard deviations are used as a proxy for the idiosyncratic risk attached to a pair. Factor regressions are performed separately for each of the 12-month formation periods.

Second, if transaction costs contribute to the persistence of price deviations, returns should be related to liquidity. As ETF liquidity is multi-faceted (see the discussions in section 2 and 3), I construct a variety of different proxies. First, I compute the natural logarithm of pair-average total net assets, recorded on the last day of the preceding formation period. Second, I use the median of daily bid-ask spreads as a proxy for transaction costs involved with normal-sized quantities. Third, I compute Amihud illiquidity ratios (Amihud, 2002) to measure the price impact costs associated with larger transactions. Fourth, I employ both the median of daily turnover ratios and the natural logarithm of average daily trading volumes as an overall proxy for trading activity. The aforementioned variables primarily capture secondary market liquidity. However, ETF shares are also tradable through the primary market by using the creation/redemption mechanism. As discussed by Broman and Shum (2016), the relation between primary market and secondary market liquidity measures is less straightforward. For example, if newly created shares are not loaded off in the secondary market, share creations lead to a decrease in turnover ratios, as the denominator (shares outstanding) grows, but the nominator is unaffected. In the opposite case of share redemptions, turnover ratios increase. Similarly, trading volumes only reflect share creation activity if newly created share are regularly sold in the secondary market. Thus, to capture all different facets of ETF liquidity, it is necessary to proxy for primary market liquidity as well, which ultimately depends on the liquidity of the underlying assets. Unfortunately, underlying liquidity data is difficult to obtain from common databases, especially for more exotic indices. Hence, I follow Broman and Shum (2016) and proxy primary market liquidity by the share creation/redemption activity, computed as

$$PrimActivity_{i,T} = \log\left(1 + \frac{1}{T}\sum_{t=1}^{T} \frac{|SHR_{i,t} - SHR_{i,t-1}|}{SHR_{i,t-1}}\right), \tag{3}$$

where $SHR_{i,t}$ is the number of shares of ETF i outstanding on day t. All of the aforementioned liquidity variables are first measured individually for the paired-up funds over the preceding 12-month formation periods, and the individual measures are then used to compute equally-weighted pair-level averages.

Table 10 shows how arbitrage profits relate to cross-sectional differences in transaction and holding costs. For this purpose, I split all trades into the highest and lowest quintiles with respect to the proxies discussed above. The results show that returns are indeed significantly related to a number of different arbitrage costs. First, pairs carrying relatively high idiosyncratic risk exhibit substantially higher returns than pairs that are only slightly exposed to idiosyncratic risk, with the latter even suffering losses on average. The difference of 206 bps on average is highly significant in both economic and statistical terms. Thus, though convergence risk is quite limited compared to other contexts (such as cross-listings, e.g. Gagnon and Karolyi, 2010), relative ETF price gaps are related to holding costs as proxied by idiosyncratic risk.

[insert Table 10 here]

Second, returns are also significantly related to four out of six liquidity proxies. Mispricings are more profitable among pairs trading at larger bid-ask spreads and for pairs with higher market impact costs as measured according to Amihud (2002). With respect to trading activity measures, evidence is rather mixed: while returns do not significantly differ for pairs with high and low dollar trading volume, they are at least slightly related to turnover. Besides, pairs with little share creation/redemption activity generate a 119 bps and thus considerably larger return than pairs with a rather active primary market. Finally, though having the predicted sign, there is no statistically significant difference in the profitability of mispricings in small and large funds.

[insert Table 11 here]

To better isolate the effect of primary market liquidity from the impact of other arbitrage costs, I also perform two-way sorts. Specifically, I first sort all trades into quartiles with respect to their primary market activity. Within these four buckets, I then sort all trades again into quartiles, this time separately by one of the remaining arbitrage costs proxies discussed above. Returns were then reported separately for each group. The results are presented in Table 11 and can be summarized as follows. First, regardless of the arbitrage cost proxy used for the second sort, price gaps are consistently more profitable within pairs having relatively inactive primary markets. Return differences are in all cases significantly larger among pairs from the lowest compared to the highest primary market activity quartile, in both economic and statistical terms. It should be noted that in principal, funds can have inactive primary markets either because they are always priced efficiently, or because there are frictions impeding the creation or redemption of shares. The observation of significantly higher arbitrage profits within low-activity pairs rather supports the first implication. If arbitrage using the creation/redemption mechanism is impeded, price correction is left to secondary market trading, ultimately forcing arbitrageurs to take convergence risk (see section 3). Consequently, mispricings have to be larger among these pairs in order to compensate for the additional risk. In results unreported for brevity, I found that low-activity pairs often track indices that may at least occasionally be difficult to trade, including non-domestic equities, small caps with factor tilts, aggregate bond indices, or physically-backed precious metal funds. In the latter case, share creations literally involve a physical delivery of bullions into the vault of the fund's custodian, which especially for large transactions may be more difficult than accumulating exchange-traded underlyings.[19]

Table 11 also reveals that within each primary market activity quartile, mispricings are significantly more pronounced within fund pairs exhibiting higher bid-ask spreads, higher Amihud (2002) illiquidity, and higher idiosyncratic risk. Evidence is mixed with respect to total net assets, trading volume, and turnover. Mispricings are larger among smaller and more actively traded fund pairs. However, this observation does not hold consistently for all primary market activity quartiles likewise. For pairs with

[19] A closer look at the data also revealed that there is no substantial difference in the frequency of primary market transactions between precious metal and other ETFs, but quantities (measured in percent of the shares outstanding) are typically smaller for these funds. Besides, significantly higher average returns for low-activity funds can still be observed when only focusing on equity ETFs. The results are available upon request.

below-median primary market activity, mispricings are more pronounced among larger and more actively traded pairs. On the contrary, pairs with active primary markets tend to exhibit larger mispricings among smaller and less actively traded pairs.

Overall, I infer from these findings that idiosyncratic risk, primary market activity, and the variety other liquidity proxies used proxy for different facets of arbitrage costs and risks, i.e. reflect separately binding impediments to arbitrage that contribute to the persistence of relative ETF price gaps.

5.3.4 Within-Pair Liquidity Differences

The results so far provided insights into the relation between *pair-level* differences in arbitrage costs and profitability. However, as discussed in section 5.2, the vast majority of pairs exhibits considerable *within-pair* differences with respect to size and liquidity. If long-short positions typically consisted of long positions in illiquid ETFs and short positions in relatively liquid ETFs, arbitrage profits could be interpreted as a premium for providing liquidity (see Schultz and Shive, 2010). For this reason, Table 12 reports net-of-fee returns separately for both trades where the more liquid fund is the position's long leg and trades where the more illiquid fund forms the long leg. The results have two implications. First, at least when measuring liquidity by bid-ask spreads, Amihud illiquidity or turnover ratios, there is indeed evidence that returns are significantly larger for arbitrage positions in which the more illiquid of both funds is held long. Nevertheless, returns remain positive and significant for mispricings in which the more liquid ETF is the position's long leg, implying that arbitrage profits can only in part be construed as premium for liquidity provision. Second, as illiquid assets are typically more difficult to borrow, these findings provide additional evidence that short-sale constraints cannot explain why price gaps persist. Otherwise, returns should be larger for positions where illiquid funds are held short. The results point in the opposite direction.

[insert Table 12 here]

5.4 Time-Varying Limits to Arbitrage and Multivariate Evidence

Aside from cross-sectional differences in the *level* of liquidity, it is widely discussed in the literature that asset prices should also be linked to liquidity shocks, i.e. sudden *changes* in the level of liquidity (e.g., Campbell, Grossman, and Wang, 1993; Engelberg, Gao, and Jagannathan, 2009). In particular, Marshall, Nguyen, and Visaltanachoti (2013) show that there is a negative liquidity shock in the minutes surrounding intraday arbitrage opportunities between two S&P 500 ETFs.

Besides the time-varying liquidity of single assets, it is also well-known that time-varying market-wide impediments to arbitrage contribute to the persistence of mispricings in a variety of settings (e.g., Asness, Moksowitz, and Pedersen, 2013; Jacobs, 2015). The most widely used measures to proxy for market-level limits to arbitrage are the CBOE Volatility Index (VIX) and the TED spread (e.g., Brunnermeier, Nagel, and Pedersen, 2008), both capturing the overall availability of arbitrage capital. In the context of ETFs, it is known that premiums and discounts are related to the VIX (e.g., Petajisto, 2017; Ben-David, Franzoni, and Moussawi, 2012). Given that *relative* ETF price gaps are typically attributable to diverging premiums and discounts (see section 5.2), it is reasonable to assume that secondary market arbitrage profits are also larger in times of a high VIX. The time-series of daily pair portfolio returns and the VIX as plotted in Figure 2 are a first indication supporting this assumption.

Figure 3 provides some initial evidence that relative ETF price deviations are also related to liquidity shocks. For the typical price gap, the pair-average bid-ask spread begins to increase slightly a few days before and peaks on the day the arbitrage opportunity occurs. Subsequently, spreads typically revert to the pre-event level.

To disentangle the effect of market-wide limits to arbitrage, liquidity shocks and cross-sectional differences in arbitrage costs, I follow Engelberg, Gao, and Jagannathan (2009) and perform multivariate regressions with the returns of single arbitrage positions as unit of observation. As explanatory variables, I use a number of proxies for cross-sectional differences in limits to arbitrage from section 5.3, the VIX on the day the price gap emerges, as well as two measures proxying for liquidity shocks on that day. Specifically, I follow Engelberg, Gao, and Jagannathan (2009) and compute

(i) the change of the pair-level average bid-ask spread, measured over the 5 days preceding the price gap, minus the pair's average daily spread over the 12-month formation period preceding the current trading period,
(ii) the change of the pair-level average turnover, measured over the 5 days preceding the price gap, minus the pair's average daily turnover over the 12-month formation period preceding the current trading period.

I also include a number of commonly used control variables, such as daily factor premia on the event-day and indicator variables for year, month, day of the week and the underlying index. The results are presented in Table 13 and can be summarized as follows. First, the estimated coefficients on the proxies for cross-sectional differences in arbitrage costs largely support the results discussed in section 5.3.2. Returns are consistently related to cross-sectional differences in the level of liquidity when proxied using the pair-average bid-ask spread and turnover. Besides, the observation that arbitrage profits are negatively related to primary market activity as already discussed in section 5.3.3 even holds in a multivariate setting. However, the economic magnitude is small compared to the other liquidity proxies considered: a one standard deviation decrease in primary market activity implies a 5 to 15 bps increase in arbitrage profits. Returns are also larger for pairs with higher idiosyncratic risk, providing further evidence that though convergence risk is quite limited among ETF pairs (see section 5.2), holding costs still tend to be an important impediment to arbitrage. Nevertheless, the economic magnitude is small compared to liquidity as measured by the bid-ask spread. A one standard deviation increase in idiosyncratic risk implies a 31-44 bps rise in arbitrage profits, compared to 78 to 90 bps for the bid-ask spread. Thus, it can be inferred that in economic terms, the most important impediment contributing to the persistence of relative ETF price gaps are transaction costs.

It is hence not surprising that price deviations are not only larger among pairs with a lower level of liquidity, but with one exception also related to sudden drops in liquidity: mispricings are typically associated with an increase in bid-ask spreads and a decrease in turnover. Finally, arbitrage profits increase with market-wide limits to arbitrage as measured by the VIX on the day of divergence. Thus, price gaps tend to be larger in times of a high VIX, when the overall availability of arbitrage capital is limited.

For all regressions performed, intercepts are positive and highly significant in both statistical and economic terms. Nevertheless, the estimates strongly suggest that impediments to arbitrage are often binding, providing a rational explanation for the persistence of mispricings. Altogether, the limits to arbitrage proxies considered can explain up to 20 percent of the entire variation in arbitrage profits.

6 Conclusion

As the combination of intraday tradability and open-endedness provides a natural arbitrage channel, ETFs were long time assumed to be priced quite efficiently. However, more recent evidence shows that ETFs can and frequently do trade at premiums or discounts on their NAV. In this paper, I evaluate the pricing efficiency of ETFs by studying the profitability and determinants of an arbitrage strategy in the secondary market for ETF shares. I combine qualitative criteria and formal co-integration tests to carefully identify near-perfect substitute funds, and subsequently open long-short positions in these pairs once prices diverge by more than two historical standard deviations. The underlying notion is that if ETFs are priced efficiently, prices should not only be aligned with NAVs, but it should neither be possible to profitably arbitrage nearly identical funds against each other. While relative ETF mispricings have already been subject of discussion in previous studies (Marshall, Nguyen, and Visaltanachoti, 2012; Petajisto, 2017), my paper is the first to provide comprehensive evidence by covering a wide range of properly selected substitute-funds, accounting for trading costs, and studying the extent to which price deviations can be attributed to cross-sectional and time-varying limits to arbitrage.

I find that though price deviations amount to only about 1 percent on average, they occur frequently enough to allow for a profitable implementation of a long-short arbitrage strategy exploiting these discrepancies. On a portfolio level, such a strategy historically generated excess returns in the order of 2.5 to 4 percent per year, net of bid-ask spreads and common estimates for commissions, short rebates and margin interest. These results cast serious doubt on the efficiency of ETF prices, though the returns reported in this paper are considerably smaller than in previous studies that implemented more simple strategies and did not account for trading costs (e.g., Fulkerson, Jordan, and Riley, 2014).

Arbitrage profits can neither be explained by fundamental risk, nor do they disappear when excluding funds that may be notoriously difficult to borrow. Given that there is no convincing evidence for fundamental risk, and considering that the selected pairs trade in the same market, ETF pairs are a somewhat unique setting to examine the profitability and limits of relative-value arbitrage as a price-correcting mechanism (see also Marshall, Nguyen, and Visaltanachoti, 2012). In previously studied contexts, asset pairs either were fundamentally different (e.g. dual-class shares, Schultz and Shive, 2010) or traded in different markets, i.e. were subjected to different institutional features (e.g., dual-listed companies as in De Jong, Rosenthal, and Van Dij, 2009; cross-listings as in Gagnon and Karolyi, 2010).

However, while fundamental risk is negligible among ETF pairs, arbitrage positions are still exposed to some convergence risk. Consistent with the latter observation, I find that arbitrage profits are strongly related to cross-sectional differences in holding costs as proxied by idiosyncratic risk. Besides, I also find that price gaps are significantly larger among more illiquid pairs as measured by a variety of different liquidity proxies. In particular, returns are substantially larger for pairs with low "on-screen" liquidity as measured by the bid-ask spread, turnover, and among pairs with low primary market activity, i.e. pairs where arbitrage through the share creation/redemption mechanism is likely to be impeded. Mispricings also tend to be larger on days with sudden drops in pair-level liquidity and on days with larger market-wide impediments to arbitrage. Together, these limits to arbitrage proxies can explain up to 20 percent of the variation in arbitrage profits. Overall, the results suggest that the persistence of relative ETF mispricings can at least to a notable extent be attributed to the existence of impediments to arbitrage.

Though unable to explain the entire return variation, these findings provide a plausible justification for the persistence of mispricings. However, even if arbitrage profits could be entirely explained by limits to arbitrage, one cannot ignore the fact ETF prices are not efficient in the proper sense. The existence of arbitrage costs and risks, imposing an impediment for information diffusion, can solely explain why ETF prices do not always fully reflect all available information, but they do not allow to reconcile price deviations with the concept of market efficiency (see also Pontiff, 2006). The most intuitive implication of my research for investors is that they are well-advised to compare the prices of competing funds tracking the desired index before trading, as it is not unusual for two nearly identical funds to exhibit at considerably different prices in both statistical and economic terms. This should be a particular concern for individuals using ETFs for short-term trading rather than long-term investing, which is a behavior apparently applying to many ETF investors.[20]

Future versions of this paper could be improved by further examining the conditions prevailing on the day the mispricing occurs. What drives the observed liquidity shocks at all? In the context of stock pairs trading, Jacobs and Weber (2015) suggest that arbitrage profits primarily stem from the asynchronous

[20] See, for example, the Vanguard research paper „*ETFs: For the better or the bettor?*", released in July 2013.

reaction to information affecting both securities likewise. Another interesting extension would be to further examine the interaction between relative price gaps and the intensity index competition. Does the magnitude and frequency of price deviations decrease with the number of similar funds due to competition effects, or does the segmentation of liquidity actually lead to an increase?

References

Abreu, Dilip and Markus Brunnermeier (2002): Synchronization Risk and Delayed Arbitrage, Journal of Financial Economics 66, 341-360.

Ackert, Lucy, and Yisong Tian (2000): Arbitrage and Valuation in the Market for Standard & Poor's Depositary Receipts, Financial Management 29(3), 71-87.

Alsayed, Hamed, and Frank McGroarty (2012): Arbitrage and the Law of One Price in the Market for American Depository Receipts, Journal of International Financial Markets, Institutions and Money 22(5), 1258-1276.

Amihud, Yakov (2002): Illiquidity and Stock Returns: Cross-Section and Time-Series Effects, Journal of Financial Markets 5(1), 31-56.

Angel, James, Todd Broms, and Gary Gastineau (2016): ETF Transaction Costs Are Often Higher Than Investors Realize, Journal of Portfolio Management, 42(3), 65-75.

Asness, Clifford, Tobias Moskowitz, and Lasse Pedersen (2013): Value and Momentum Everywhere, Journal of Finance 68(3), 929-985.

Avellaneda, Marco, and Doris Dobi (2013): Price Inefficiency and Stock-Loan Rates of Leveraged ETFs, RISK Magazine (July 16, 2013), 37-40.

Barberis, Nicholas, and Richard Thaler (2003): A Survey of Behavioral Finance, in: Handbook of the Economics and Finance, Vol. 1, pp. 1053-1128.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi (2017): Exchange Traded Funds (ETFs), Annual Review of Financial Economics 9, forthcoming.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi (2017): Do ETFs Increase Volatility? NBER Working Paper.

Borkovec, Milan, and Vitaly Serbin (2013): Create or Buy: A Comparative Analysis of Liquidity and Transaction Costs for Selected US ETFs, Journal of Portfolio Management, 39(4), 118-131.

Box, Travis, Ryan Davis, and Kathleen Fuller (2017): ETF Competition and Market Quality. Working Paper.

Broman, Markus (2016): Liquidity, Style Investing, and Excess Comovement of Exchange-Traded Fund Returns, Journal of Financial Markets 30, 27-53.

Broman, Markus (2016): Local Demand, Preferred Habitats and Excess Comovement. Working Paper.

Broman, Markus, and Pauline Shum (2016): Does Liquidity Encourage Short-Term Trading? Evidence from Exchange-Traded Funds. Working Paper.

Brunnermeier, Markus, and Lasse Pedersen (2009): Market Liquidity and Funding Liquidity, Review of Financial Studies 22(6), 2201-2238.

Brunnermeier, Markus, Stefan Nagel, and Lasse Pedersen (2008): Carry Trades and Currency Crashes. NBER Working Paper.

Campbell, John, Sanford Grossman, and Jiang Wang (1993): Trading Volume and Serial Correlation in Stock Returns, Quarterly Journal of Economics 108, 905-939.

Carhart, Mark (1997): On Persistence in Mutual Fund Performance, Journal of Finance 52(1), 57-82.

Charupat, Narat, and Peter Miu (2011): The Pricing and Performance of Leveraged Exchange-Traded Funds, Journal of Banking & Finance 35(4), 966-977.

De Jong, Abe, Leonard Rosenthal and Mathijs van Dijk (2009): The Risk and Return of Arbitrage in Dual-Listed Companies, Review of Finance 13(3), 495-520.

De Long, Bradford, Andrei Shleifer, Lawrence Summers, and Robert Waldmann (1990): Noise Trader Risk in Financial Markets, Journal of Political Economy 98(4), 793-738.

Delcoure, Natalya, and Maosen Zhong (2007): On the Premiums of iShares, Journal of Empirical Finance 14(2), 168-195.

Do, Binh, and Robert Faff (2016): Cointegration and Relative Value Arbitrage. Working Paper.

Engelberg, Joseph, Pengjie Gao, and Ravi Jagannathan (2009): An Anatomy of Pairs Trading: The Role of Idiosyncratic News, Common Information and Liquidity. Working Paper.

Engle, Robert, and Clive Granger (1987): Co-integration and Error Correction: Representation, Estimation, and Testing, Econometrica 55(2), 251-276.

Engle, Robert, and Debojyoti Sarkar (2006): Premiums-Discounts and Exchange Traded Funds, Journal of Derivatives, 27-45.

Fama, Eugene, and Kenneth French (1992): The Cross-Section of Expected Stock Returns, Journal of Finance 47(2), 427-465.

Fama, Eugene, and Kenneth French (2015): A Five-Factor Asset Pricing Model, Journal of Financial Economics 116(1), 1-22.

Froot, Kenneth, and Emil Dabora (1999): How Are Stock Prices Affected by the Location of Trade?, Journal of Financial Economics 53, 189-216.

Fulkerson, Jon, Susan Jordan, and Denver Travis (2017): Bond ETF Arbitrage Strategies and Daily Cash Flow, Journal of Fixed Income 27(1), 49-65.

Fulkerson, Jon, Susan Jordan, and Timothy Riley (2014): Predictability in Bond ETF Returns, Journal of Fixed Income 23(3), 50-63.

Gagnon, Louis, and George Karolyi (2010): Multi-Market Trading and Arbitrage, Journal of Financial Economics 97(1), 53-80.

Gatev, Evan, William Goetzmann, and Geert Rouwenhorst (2006): Pairs Trading: Performance of a Relative Value Arbitrage Rule, Review of Financial Studies 19(3), 797-827.

Glosten, Lawrence, Suresh Nallareddy, and Yuan Zo (2017): ETF Activity and Informational Efficiency of Underlying Securities. Columbia Business School Working Paper.

Hilliard, Jitka (2014): Premiums and Discounts in ETFs: An Analysis of the Arbitrage Mechanism in Domestic and International funds, Global Finance Journal 25(2), 90-107.

Jacobs, Heiko and Martin Weber (2015): On the Determinants of Pairs Trading Profitability, Journal of Financial Markets 23, 75-97.

Jacobs, Heiko (2015): What Explains the Dynamics of 100 Anomalies?, Journal of Banking & Finance 57, 65-85.

Jegadeesh, Narasimhan (1990): Evidence of Predictable Behavior of Security Returns, Journal of Finance 45, 881-898.

Jegadeesh, Narasimhan, and Sheridan Titman (1990): Short-Horizon Return Reversals and the Bid-Ask Spread, Journal of Financial Intermediation 4(2), 116-132.

Jiang, Xinxin, and Stanley Peterburgsky (2017): Investment Performance of Shorted Leveraged ETF Pairs, Applied Economics 49, 4410-4427.

Johnson, Timothy (2008): Volume, Liquidity, and Liquidity Risk, Journal of Financial Economics 87(2), 388-417.

Levy, Ariel, and Offer Lieberman (2012): Overreaction of Country ETFs to US Market Returns: Intraday vs. Daily Horizons, Journal of Banking & Finance 37(5), 1412-1421.

Madhavan, Ananth, and Aleksander Sobczyk (2016): Price Dynamics and Liquidity of Exchange-Traded Funds, Journal of Investment Management 14(2), 1-17.

Marshall, Ben, Nhut Nguyen, and Nuttawat Visaltanachoti (2013): ETF Arbitrage: Intraday Evidence, Journal of Banking & Finance, 37, p. 3486-3498.

Marshall, Ben, Nhut Nguyen, and Nuttawat Visaltanachoti (2017): Do Liquidity Proxies Measure Liquidity Accurately in ETFs? Working Paper.

Mitchell, Mark, Pulvino, Todd and Stafford, Erik (2002): Limited Arbitrage in Equity Markets, Journal of Finance, 57(2), p. 551-584.

Petajisto, Antti (2017): Inefficiencies in the Pricing of Exchange-Traded Funds, Financial Analysts Journal 73(1), 24-54.

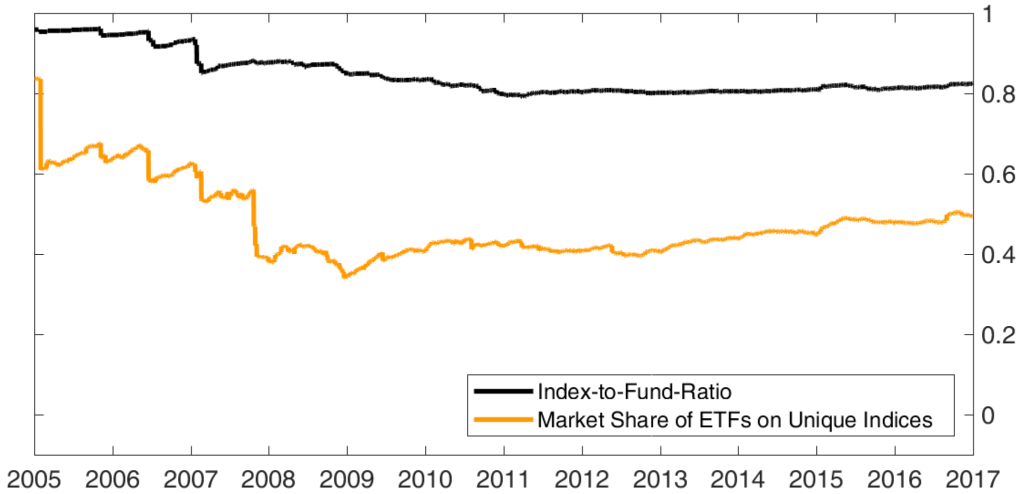
Pontiff, Jeffrey (2006): Costly Arbitrage and the Myth of Idiosyncratic Risk, Journal of Accounting and Economics 42(1), 35-52.

Schultz, Paul and Shive, Sophie (2010): Mispricing of Dual-Class Shares: Profit Opportunities, Arbitrage, and Trading, Journal of Financial Economics, 98, p. 524-549.

Stratmann, Thomas, and John Welborn (2013): Exchange-Traded Funds, Fails-to-Deliver, and Market Volatility. Working Paper.

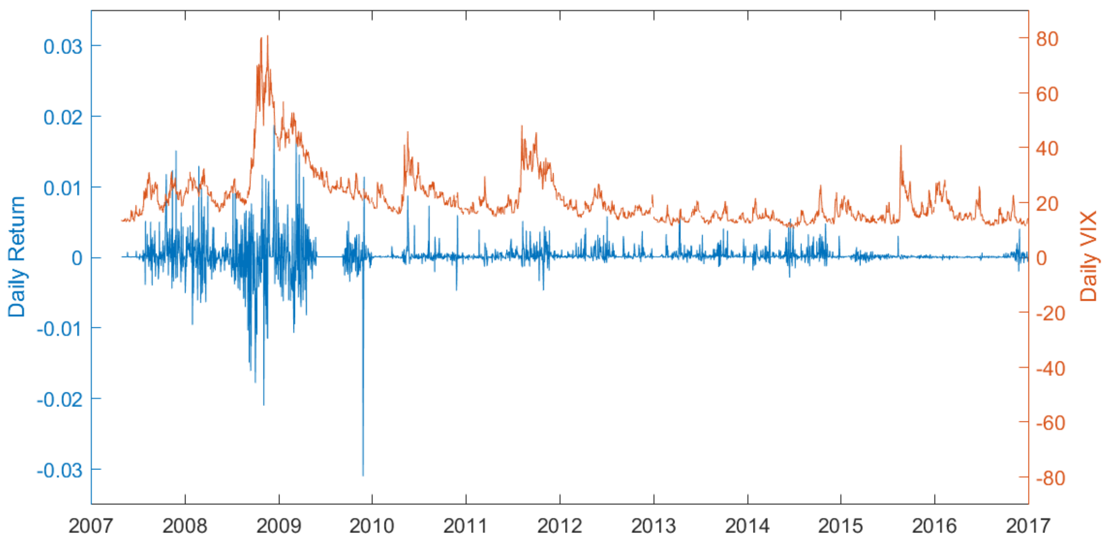
Appendix 1 – Figures

Figure 1: Index Homogeneity in the US ETF Market (2005-2016)



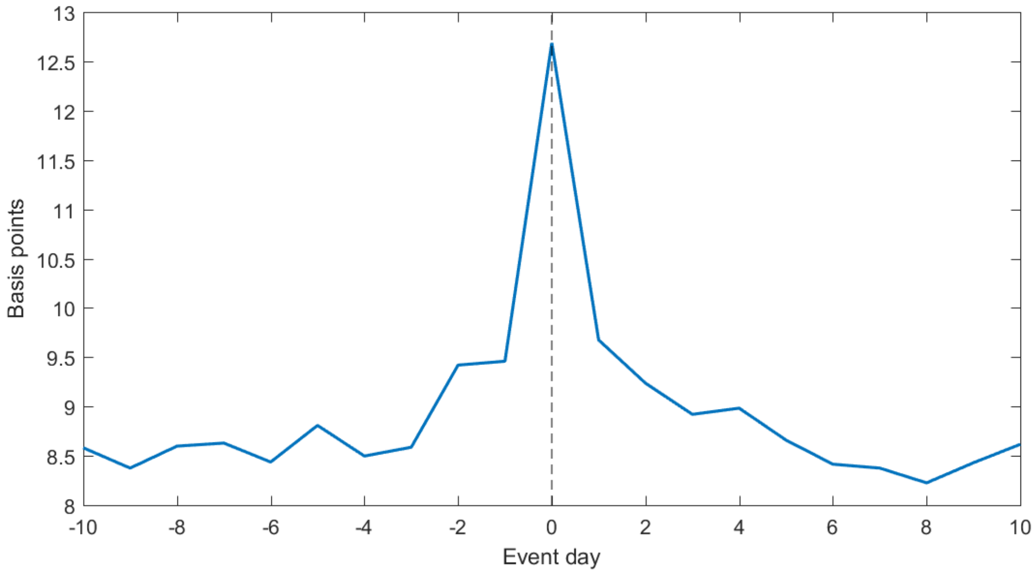
This figure illustrates the homogeneity of the US ETF market over time. More specifically, it plots the (i) ratio of unique investable indices to investable ETFs, as well as (ii) the market share of unique ETFs, i.e. funds for which there is not a single competing fund tracking the same index.

Figure 2: Daily Portfolio Returns and VIX Over Time (2007-2016)



This figure plots daily net-of-fee portfolio returns (measured as return on employed capital) against daily VIX levels. Daily VIX series were obtained from the St. Louis Fed Database (FRED Economic Data). Portfolio returns are computed as outlined in section 4.2.3.

Figure 3: Bid-Ask Spread Around the Event-Day



This figure plots the pair-level bid-ask spread for the 20 trading days around the day of the mispricing, i.e. around the day where positions are initiated. Spreads were computed by first taking the equally weighted pair-level average spreads for the 20 days surrounding a specific mispricing in that pair. The figures plotted were then obtained by separately computing the median for each of the event-window days across all mispricings.

Appendix 2 – Tables

Table 1: Sample Characteristics

	Mean	StdDev	Percentiles						
			Min	5	25	50	75	95	Max
Panel A: 2007-2016									
Net assets ($ millions)	1,405.8	7,365.3	0.2	1.8	11.8	77.5	489.9	6,250.6	224,820.2
Daily volume ($ millions)	33.3	534.5	0.0	0.0	0.1	0.5	2.8	61.3	23,816.1
Daily turnover (%)	59.8	1,833.1	0.0	0.2	0.8	1.3	2.2	13.5	64,745.9
Bid-ask spread (bps)	44	75	1	4	11	22	50	133	843
Premium (bps)	2	75	-1,985	-25	-1	2	12	43	517
Premium volatility (bps)	59	59	0	5	21	46	79	154	774
Panel B: 2007-2011									
Net assets ($ millions)	814.7	4,175.6	0.5	2.5	10.3	50.1	287.7	3,104.8	95,397.4
Daily volume ($ millions)	67.0	825.9	0.0	0.0	0.2	0.9	5.1	133.3	26,211.3
Daily turnover (%)	7.6	41.0	0.0	0.5	1.0	1.6	3.0	15.8	625.5
Bid-ask spread (bps)	35	64	1	4	10	20	34	101	792
Premium (bps)	5	100	-1,879	-27	-1	3	16	62	1,041
Premium volatility (bps)	84	71	0	16	38	65	107	209	774

This table shows cross-sectional ETF characteristics for the full sample period (Panel A) as well as the first five years of the sample (Panel B). Net assets are recorded at the end of the respective period. Daily volume, turnover and premium are computed as means throughout the period, whereas bid-ask spread is the time-series median and premium volatility is the time-series volatility.

Table 2: Descriptive Pair Statistics

Panel A: Pair Frequency		
Number of tested pairs	4,476	
Number of selected pairs	3,677	
Average number of tested pairs	39	
Average number of selected pairs	32	
Panel B: Characteristics of Paired-Up Funds		
Net assets ($ mil)	7,386	(636)
Bid-ask spread (bps)	13	(5)
Volume ($ mil)	501	(4)
Premium (bps)	2	(1)
Premium volatility (bps)	33	(17)
Premium correlation	0.39	(0.30)
Pairs with significant correlation (%)	65	
Panel C: Co-Integration Statistics for Selected Pairs		
Co-integrating beta (slope coefficient)	1.62	(0.98)
Constant ($)	0.86	(0.09)
Residual volatility ($)	0.47	(0.21)
Days where mean is crossed (%)	24	
Co-integrated in trading period (%)	75	

This table shows the frequency of pairs, liquidity and premium characteristics of the funds selected in pair, and pairwise co-integration statistics, pooled across all trading cycles. Panel A reports the frequency of pairs. Panel B shows cross-sectional characteristics for all funds selected in at least one pair. Net assets are recorded at the end of the respective formation periods. Daily volume and premium are computed as fund-level means throughout the respective formation periods, whereas bid-ask spread is the time-series median and premium volatility is the time-series volatility. Premium correlation is the pair-wise Pearson correlation coefficient for the premium/discount series over the formation period. "Pairs with significant correlation" refers to the share of pairs, whose premium/discount series are significantly correlated at the 5-percent-level. Panel C shows the estimated coefficients of the co-integration regression (see equation (1)), estimated over the respective formation periods. "Co-Integrated in trading period" refers to the share of pairs that remain statistically co-integrated during the trading period, i.e. where the hypothesis of co-integration cannot be rejected at the 1-percent-level in both the formation and subsequent trading period. In both Panel B and C, means are reported. Values in parentheses represent median values.

Table 3: Monthly Return Distribution and Performance Measures (2007-2016)

	Gross-of-fee Returns		Net-of-fee Returns		Market
	ROCC	ROEC	ROCC	ROEC	
Panel A: Return Distribution					
Average Return	0.0035	0.0048	0.0027	0.0036	0.0067
t-Statistic	3.5701 ***	3.7737 ***	3.0077 ***	2.9883 ***	1.5723 *
Average Excess Return	0.0031	0.0043	0.0022	0.0032	0.0063
t-Statistic	3.3398 ***	3.5362 ***	2.6575 ***	2.6834 ***	1.4669 *
Return Distribution					
Standard Deviation	0.0072	0.0105	0.0068	0.0101	0.0458
Median	0.0018	0.0030	0.0013	0.0022	0.0124
Skewness	0.4500	-0.6511	0.2734	-0.7956	-0.6537
Kurtosis	9.4845	9.3043	10.1580	9.9862	4.2731
Min	-0.0284	-0.0440	-0.0289	-0.0449	-0.1715
Max	0.0335	0.0403	0.0307	0.0388	0.1135
Observations < 0	0.0948	0.0948	0.2241	0.2155	0.3879
Value-at-Risk	-0.0064	-0.0151	-0.0066	-0.0161	-0.0812
Conditional Value-at-Risk	-0.0148	-0.0275	-0.0152	-0.0284	-0.1054
Panel B: Risk-Adjusted Performance					
Sharpe Ratio	0.4255	0.4172	0.3299	0.3127	0.1372
Excess Return on VaR	0.4793	0.2882	0.3383	0.1967	0.0771
Excess Return on CVaR	0.2063	0.1582	0.1474	0.1113	0.0594

Summary statistics for the monthly returns on the ETF arbitrage strategy and for monthly market returns. Market returns are obtained from the Kenneth French Data Library. "ROCC" is return on committed capital, whereas "ROEC" is return on employed capital. Both value-at-risk and conditional value-at-risk are computed using a 0.95 confidence level. ***, **, * means statistically significant on a 1 percent, 5 percent and 10 percent level according to the t-statistic, respectively. t-statistics reported are computed using Newey-West standard errors with six-lag correction.

Table 4: Systematic Risk for Net-of-Fee Portfolio Returns (2007-2016)

	Return on Committed Capital (ROCC)			Return on Employed Capital (ROEC)		
	(1)	(2)	(3)	(1)	(2)	(3)
Factors						
Intercept	0.0021 **	0.0021 **	0.0023 ***	0.0030 **	0.0030 **	0.0033 ***
Market	0.0396	0.0325	0.0323	0.0500	0.0399	0.0355
SMB	-0.0032	-0.0018	0.0056	0.0311	0.0331	0.0441
HML	0.0050	-0.0142	0.0379 *	0.0088	-0.0183	0.0621 *
MOM		-0.0258 ***			-0.0366 ***	
RMW			0.0133			0.0134
CMA			-0.1049 *			-0.1865 **
R-Squared	0.07	0.10	0.11	0.07	0.10	0.12

This table provides alphas and systematic risk exposures for the ETF arbitrage strategy, based on monthly net-of-fee returns. Specification (1) corresponds to Fama-French three-factor model (Fama and French, 1992), whereas (2) refers to the Carhart four-factor model (Carhart, 1997). Specification (3) is the Fama-French five-factor model (Fama and French, 2015). All factor premiums are sourced from the Kenneth French Data Library. ***, **, * means statistically significant on a 1 percent, 5 percent and 10 percent level according to the t-statistic, respectively. t-statistics are computed using the Newey-West standard errors with six-lags correction.

Table 5: Pair Portfolio Composition

Panel A: Size and Liquidity	
Weight of funds in top three size deciles	0.64
Weight of funds in top five size deciles	0.85
Weight of funds from different size deciles	0.74
Size decile difference	2.77
Weight of funds in lowest three spread deciles	0.66
Weight of funds in lowest five spread deciles	0.83
Weight of funds from different spread deciles	0.76
Spread decile difference	2.47
Panel B: Asset Class and Index Type	
Weight of equity fund pairs	0.79
Weight of commodity fund pairs	0.11
Weight of fixed income fund pairs	0.10
Weight of "smart beta" fund pairs	0.39
Panel C: Other	
Weight of same-brand pairs	0.02
Weight of pairs with same replication method	0.62

This table provides information on the characteristics of pairs portfolios, averaged across all trading periods. Panel A shows the average weights of funds that meet certain liquidity and size characteristics. Panel B shows the average weight of pairs according to their underlying asset class and index type (cap-weighted or "smart beta", i.e. factor-weighted). Panel C show the average weight of pairs, in which both ETFs are issued by the same fund sponsor or where both use the same index replication methodology.

Table 6: Price and NAV Deviations on Event Days

		Open	Close	Diff	*p*-Val	
Relative price deviation	Mean	0.0120	-0.0005	-0.0126	0.0001	***
	Median	0.0080	-0.0004	-0.0084		
Relative NAV deviation	Mean	0.0016	0.0015	-0.0001	0.2323	
	Median	0.0006	0.0007	0.0001		
Premium difference	Mean	0.0095	-0.0031	-0.0126	0.0001	***
	Median	0.0061	-0.0007	-0.0068		

This table reports relative price and NAV deviations as well as premium differences for both the days positions are established and the days positions are closed, measured across all trades triggered between 2007 and 2016. Relative price (NAV) deviations are defined as percentage difference between the mid-quote prices (NAVs) of the short and long leg. Premium differences are defined as premium of the short leg minus the premium of the long leg. *p*-Values refer to paired-sample *t*-tests with heteroskedasticity-adjusted standard errors. ***, **, * means statistically significant on a 1 percent, 5 percent and 10 percent level according to *t*-tests with standard errors clustered for the day where the mispricing occurs and adjusted for heteroskedasticity.

Table 7: Trading Statistics and Position Return Distribution

Total number of total trades		5,308	Mean return (gross of fees)	0.0126	***
Total number of converged trades		4,353	Net-of-fee returns		
Convergence probability		0.82	Mean	0.0100	***
			Mean: converged trades	0.0128	***
			Mean: unconverged trades	-0.0029	***
Duration of converged trades (days)	Mean	9.5	Standard deviation	0.0185	
	Median	4.0	Min	-0.3502	
Number of pairs traded per period	Mean	19	P01	-0.0145	
	Median	17	P10	-0.0023	
Share of pairs traded per period	Mean	0.59	P25	-0.0004	
	Median	0.65	P50	0.0063	
Number of trades per pair	Mean	1.4	P75	0.0190	
	Median	1.0	P90	0.0303	
Number of roundtrips per traded pair	Mean	2.0	P99	0.0529	
	Median	1.0	Max	0.2048	

This table shows the frequency of mispricings (left side) and the distribution of single position returns (right side), pooled across all trading cycles. ***, **, * means statistically significant on a 1 percent, 5 percent and 10 percent level according to *t*-tests with standard errors clustered for the day where the mispricing occurs and adjusted for heteroskedasticity.

Table 8: Differences in Replication Methods and Convergence Risk

	Same Replication		Diff. Replication		Difference	
Number of trades	2,935		2,373		562	
Mean return	0.0117	***	0.0074	***	0.0043	***
Mean return (converged)	0.0155	***	0.0090	***	0.0065	***
Mean return (unconverged)	-0.0044	***	-0.0012	*	-0.0032	*
Standard Deviation	0.0215		0.0134		0.0081	###
Convergence probability	0.8061		0.8373		-0.0312	
Duration of converged trades (days)	14.11		14.96		-0.85	

This table reports the frequency of mispricings as well as their profitability and risk for both pairs of funds employing the same and explicitly different replication methodologies. Returns are single position net-of-fee returns. "Avg. maximum draw-down" measures the maximum relative decline from the peak position value over the position's holding period, averaged across all trades. ***, **, * means statistically significant on a 1 percent, 5 percent and 10 percent level according to *t*-tests with standard errors clustered for the day where the mispricing occurs and adjusted for heteroskedasticity. ###, ##, # means statistically significant on a 1 percent, 5 percent and 10 percent level according to the two-sample *F*-test for equal variances.

Table 9: Single Position Returns and Short-Selling Constraints

	N	(%)	Return					
			Total		Long Leg		Short Leg	
Panel A: Leveraged Funds								
Pairs of leveraged funds	35	1	0.0134	***	-0.0108		0.0244	*
Pairs of non-leveraged equivalents	1,289	24	0.0096	***	0.0149	***	-0.0050	***
Return difference			0.0038	*	-0.0257	***	0.0295	***
Panel B: Exchange-Traded Notes								
Pairs of ETNs	78	1	0.0071	***	0.0175	***	-0.0102	***
Pairs of equivalent ETFs	1,102	21	0.0100	***	0.0136	***	-0.0034	***
Return difference			-0.0029	**	0.0040		-0.0069	*
Panel C: Total								
All pairs	5,308	100	0.0100	***	0.0116	***	-0.0016	***
Without leveraged funds and ETNs	5,211	98	0.0099	***	0.0115	***	-0.0017	***
			0.0063		*0.0093*		*0.0008*	
Difference in average returns			-0.0001		-0.0001		0.0001	

This table shows the frequency and profitability of mispricings for different subsamples of pairs. For each subsample, returns are separately reported for both the long and short leg of the position. Returns are single position net-of-fee returns. Panel A shows leveraged and inverse fund pairs (for brevity grouped under "leveraged funds") versus equivalent unleveraged pairs, i.e. pairs tracking the same indices as those covered by leveraged/inverse pairs, but without leverage. Panel B contrasts pairs involving at least one Exchange-Traded Note to equivalent pairs made up entirely of Exchange-Traded Funds, i.e. ETF pairs tracking the same indices to those covered by pairs involving ETNs. Panel C shows how the results change when excluding leveraged/inverse pairs and pairs involving at least one ETN. Values reported in italics are medians. ***, **, * means statistically significant on a 1 percent, 5 percent and 10 percent level according to *t*-tests with standard errors clustered for the day where the mispricing occurs and adjusted for heteroskedasticity.

Table 10: Single Position Returns and Cross-Sectional Differences in Arbitrage Costs

						Difference	
High idiosyncratic risk	0.0194	***	Low idiosyncratic risk	-0.0011	***	0.0206	***
	0.0214			*-0.0013*		0.0228	
Low net assets	0.0101	***	High net assets	0.0095	***	0.0006	
	0.0100			*0.0036*		0.0063	
High bid-ask spread	0.0129	***	Low bid-ask spread	0.0055	***	0.0073	***
	0.0104			*0.0005*		0.0099	
High Amihud ratio	0.0214	***	Low Amihud ratio	0.0058	***	0.0156	***
	0.0196			*0.0017*		0.0179	
Low turnover ratio	0.0103	***	High turnover ratio	0.0086	***	0.0016	**
	0.0083			*0.0034*		0.0048	
Low trading volume	0.0100	***	High trading volume	0.0100	***	0.0000	
	0.0093			*0.0044*		0.0050	
Low primary market activity	0.0195	***	High primary market activity	0.0076	***	0.0119	***
	0.0198			*0.0035*		0.0163	

This table reports average returns conditioned on subsamples formed based on seven different arbitrage cost proxies, whereas "low" refers to the lowest and "high" to the highest quintile with regard to the respective measure. Numbers written in italics are median values. *Idiosyncratic risk* is computed as residual volatility of a time-series regression, where daily pair-wise return differences are regressed on the corresponding Fama and French (1992) factors. *Amihud ratio* is computed as in Amihud (2002). *Primary market activity* is defined as in equation (3). All variables are estimated using daily data over the preceding formation period, except for total net assets, which are recorded on the last day of the formation period. ***, **, * means statistically significant on a 1 percent, 5 percent and 10 percent level according to *t*-tests with standard errors clustered for the day where the mispricing occurs and adjusted for heteroskedasticity.

Table 11: Two-Way Sorts By Primary Market Activity and Other Arbitrage Cost Proxies

Primary Market Activity	**Low**	**2**	**3**	**High**	**All**	**High minus low**	
Panel A: By Bid-Ask Spread Quartiles							
Low	0.0165	0.0047	0.0059	0.0047	0.0069	0.0118	***
2	0.0118	0.0073	0.0078	0.0070	0.0098	0.0049	***
3	0.0142	0.0085	0.0095	0.0109	0.0104	0.0034	***
High	0.0311	0.0062	0.0081	0.0075	0.0119	0.0236	***
High minus low spread	0.0146	0.0015	0.0022	0.0028	0.0050		
	***	*	**	**	***		
Panel B: By Amihud Quartiles							
Low	0.0162	0.0037	0.0033	0.0018	0.0064	0.0145	***
2	0.0084	0.0029	0.0024	0.0050	0.0041	0.0034	***
3	0.0216	0.0023	0.0086	0.0068	0.0078	0.0148	***
High	0.0275	0.0178	0.0171	0.0164	0.0207	0.0111	***
High minus low Amihud	0.0112	0.0141	0.0138	0.0146	0.0143		
	***	***	***	***	***		
Panel C: By Idionsyncratic Risk Quartiles							
Low	0.0067	-0.0019	-0.0007	-0.0002	-0.0008	0.0070	***
2	0.0162	0.0025	0.0055	0.0076	0.0088	0.0086	***
3	0.0264	0.0157	0.0153	0.0145	0.0141	0.0119	***
High	0.0243	0.0107	0.0113	0.0081	0.0184	0.0162	***
High minus low idio risk	0.0176	0.0126	0.0120	0.0083	0.0192		
	***	***	***	***	***		

This table presents average net-of-fee returns of single positions, sorted by primary market activity as defined in equation (3) and six different arbitrage cost proxies, namely bid-ask spreads, Amihud (2002) illiquidity ratios, idiosyncratic risk, total net assets, trading volumes, and turnover ratios. Specifically, all trades are first sorted into quartiles with respect to their primary market activity. Within these four groups, trades are then sorted by pair-level averages of the aforementioned proxies. All measures are computed as in Table 10. "High minus low" refers to differences in average returns of trades in the highest quartile and trades in the lowest quartile. ***, **, * means statistically significant on a 1 percent, 5 percent and 10 percent level according to *t*-tests with standard errors clustered for the day where the mispricing occurs and adjusted for heteroskedasticity.

Table 11 (continued)

Primary Market Activity	Low	2	3	High	All	High minus low	
Panel D: By Size (Net Assets)							
Low	0.0136	0.0061	0.0082	0.0098	0.0080	0.0038	***
2	0.0070	0.0036	0.0070	0.0067	0.0075	0.0003	
3	0.0150	0.0045	0.0028	0.0019	0.0068	0.0131	***
High	0.0145	0.0036	0.0021	0.0007	0.0053	0.0138	***
High minus low size	0.0009	-0.0025	-0.0062	-0.0091	-0.0027		
		***	***	***	***		
Panel E: By Volume							
Low	0.0118	0.0040	0.0085	0.0103	0.0072	0.0015	*
2	0.0100	0.0040	0.0047	0.0054	0.0077	0.0045	***
3	0.0103	0.0053	0.0047	0.0034	0.0066	0.0070	***
High	0.0179	0.0045	0.0023	0.0000	0.0060	0.0179	***
High minus low volume	0.0061	0.0005	-0.0062	-0.0103	-0.0012		
	***		***	***	***		
Panel C: By Turnover							
Low	0.0135	0.0031	0.0063	0.0056	0.0074	0.0080	***
2	0.0113	0.0024	0.0043	0.0089	0.0055	0.0023	***
3	0.0063	0.0066	0.0047	0.0045	0.0074	0.0019	***
High	0.0189	0.0057	0.0049	0.0000	0.0064	0.0189	***
High minus low turnover	0.0054	0.0027	-0.0015	-0.0055	-0.0010		
	***	***	**	***	**		
All trades	0.0125	0.0045	0.0051	0.0048	0.0069	0.0077	***

This table presents average net-of-fee returns of single positions, sorted by primary market activity as defined in equation (3) and six different arbitrage cost proxies, namely bid-ask spreads, Amihud (2002) illiquidity ratios, idiosyncratic risk, total net assets, trading volumes, and turnover ratios. Specifically, all trades are first sorted into quartiles with respect to their primary market activity. Within these four groups, trades are then sorted by pair-level averages of the aforementioned proxies. All measures are computed as in Table 10. "High minus low" refers to differences in average returns of trades in the highest quartile and trades in the lowest quartile. ***, **, * means statistically significant on a 1 percent, 5 percent and 10 percent level according to *t*-tests with standard errors clustered for the day where the mispricing occurs and adjusted for heteroskedasticity.

Table 12: Liquidity Differences and Arbitrage Profitability

		By Bid-Ask Spread		By Amihud Ratio		By Turnover		By Primary Market Activity	
Long liquid ETF	N	2,818		2,711		2,591		2,519	
	Mean	0.0081	***	0.0089	***	0.0096	***	0.0105	***
Long illiquid ETF	N	2,490		2,597		2,367		2,439	
	Mean	0.0116	***	0.0106	***	0.0106	***	0.0097	***
Liquid minus illiquid	Difference	-0.0035	***	-0.0017	***	-0.0010	**	0.0008	

This table presents average net-of-fee returns of single positions separately for trades where the more liquid or the more illiquid of both ETFs is held long. Liquidity is either measured by bid-ask spreads, Amihud illiquidity ratios, turnover, or primary market activity. Liquidity measures are computed as in Table 10. “Liquid minus illiquid” refers to differences in average returns between both types of mispricings. ***, **, * means statistically significant on a 1 percent, 5 percent and 10 percent level according to *t*-tests with standard errors clustered for the day where the mispricing occurs and adjusted for heteroskedasticity.

Table 13: Cross-Sectional Regressions of Position Returns on Arbitrage Cost Proxies

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Intercept	0.0089 ***	0.0083 ***	0.0084 ***	0.0079 ***	0.0085 ***	0.0084 ***	0.0076 ***
	18.66	*19.55*	*19.18*	*19.81*	*18.40*	*19.32*	*20.84*
Bid-ask spread		0.0090 ***		0.0092 ***		0.0090 ***	0.0078 ***
		8.58		*8.64*		*8.40*	*7.38*
Turnover	-0.0023 ***	-0.0011 ***	-0.0026 ***	-0.0013 ***	-0.0023 ***	-0.0011 ***	-0.0018 ***
	-6.72	*-3.30*	*-7.40*	*-3.75*	*-6.46*	*-3.16*	*-5.71*
Amihud ratio	0.0011 **	-0.0008 **	0.0012 ***	-0.0007 **	0.0011 **	-0.0007 **	-0.0004
	2.44	*-2.34*	*2.62*	*-2.13*	*2.35*	*-2.13*	*-1.30*
PrimActivity	-0.0011 ***	-0.0005 **	-0.0015 ***	-0.0009 ***	-0.0011 ***	-0.0005 **	-0.0013 ***
	-3.14	*-2.09*	*-4.05*	*-3.56*	*-3.09*	*-2.20*	*-3.62*
Idiosyncratic risk	0.0043 ***		0.0044 ***		0.0041 **		0.0031 **
	2.69		*2.75*		*2.53*		*2.14*
VIX			0.0023 ***	0.0022 ***			0.0026 ***
			3.50	*3.16*			*3.96*
ΔSpread					0.0007 *	0.0015 ***	0.0012 ***
					1.68	*4.65*	*3.02*
ΔTurnover					-0.0007	-0.0008 *	-0.0007 *
					-1.45	*-1.95*	*-1.66*
Adjusted R^2	0.1286	0.1626	0.1451	0.1778	0.1326	0.1326	0.1935

This table provides results from cross-sectional regressions, where single position net-of-fee returns are regressed on a number of proxies for cross-sectional differences in arbitrage costs, as well as the volatility index (VIX) and the TED spread on the day the position is initiated. Cross-sectional arbitrage cost proxies include the bid-ask spread, the Amihud illiquidity ratio (Amihud, 2002), primary market activity (PrimActivity) as measured in equation 3, and idiosyncratic volatility. These variables are computed as in Table 9. Daily VIX and TED spread data were downloaded from the St. Louis Fed Database (FRED Economic Data). As control variables, I included daily factor premia in line with Fama and French (1992) measured on the day where the position is opened and indicator variables for year, month, day of the week and index underlying the respective fund pair. Factor premiums were obtained from the Kennneth French Data Library. Explanatory variables are standardized to have zero mean and unit variance. ***, **, * means statistically significant on a 1 percent, 5 percent and 10 percent level according to *t*-tests with standard errors clustered for the day where the mispricing occurs and adjusted for heteroskedasticity.