# Excess Comovement and Limits-to-Arbitrage: Evidence from Exchange-Traded Funds

Markus S. Broman*

First draft: November 28, 2012

This draft: May 20, 2013

## Abstract

The prices of Exchange-Traded Funds can deviate from their Net Asset Values by the magnitude of arbitrage costs, such as transaction and holding costs. Despite the explicit "in-kind" arbitrage mechanism that exists for ETFs, I find that ETF mispricing mean-reverts only partially; moreover, mispricing increases not only with proxies for aggregate market illiquidity, but also for ETFs with high arbitrage costs, particularly during the financial crisis. As long as mispricing remains within the arbitrage costs boundaries, there is limited pressure to correct for any price discrepancies, which may leave the ETF partially exposed to a common factor. Consistent with this prediction and with recent theoretical work on the effects of commonality in trading patterns, I find strong evidence of comovement among *excess* ETF returns (i.e. after correcting for movements in the fundamentals, the NAV) within three investment *styles* – size, value-growth and ETF liquidity.

**JEL Classification:** G10, G12, G14, G15, G20, G23

**Keywords:** Excess Comovement, ETF, mispricing, limits to arbitrage, style investing.

---

* Finance Department, Schulich School of Business, York University, 4700 Keele St., Toronto, Ontario M3J 1P3, Tel: (416) 736-2100 ext. 44655, e-mail: mbroman10@schulich.yorku.ca. I wish to thank Pauline Shum, Yisong Tian, Kee-Hong Bae and Mark Kamstra for discussions and suggestions, as well as seminar participants at York University, Schulich School of Business.

# 1 Introduction

In frictionless markets with rational investors, the market price of an Exchange-Traded Fund (ETF) will equal its fundamental price (the Net Asset Value), and any comovement in ETF returns must be due to comovement in fundamentals. However, in economies with frictions or with irrational investors, and in which there are *limits to arbitrage*, comovement in ETF returns may be partially delinked from fundamentals giving rise to what is known as *excess comovement* (see Barberis and Shleifer (2003); Barberis, Shleifer and Wurgler (2005) and Greenwood (2008)). The two main non-fundamentals based explanations of excess comovement are articulated in Barberis, Shleifer and Wurgler (2005): The first is related to *correlated investor demand* (e.g. when ETF investors engage in style investing or habitat formation), while the second arises from differences among securities in their rate of *information diffusion* (e.g. ETFs incorporate information faster than the underlying portfolio).

A unique feature of the ETF market is the "in-kind" share creation/redemption process, an *explicit* arbitrage mechanism, which allows select investors (known as Authorized Participants) to trade on price differentials between the ETF and the NAV by *exchanging* (or *receiving*) a portfolio of securities similar to the ETF's benchmark index for new ETF shares. Arbitrage is, however, *limited* by two types of costs: transaction costs (such as bid-ask spreads, market-impact costs and creation/redemption fees) and holding costs (delays in the execution of the arbitrage trade). Moreover, these arbitrage costs are expected to be magnified by the large size and relative infrequency of typical creation/redemption transactions. As long as mispricing remains *within* the arbitrage costs boundaries, there is limited pressure to correct for any price discrepancies, which may leave the ETF partially exposed to a common, non-fundamental, factor.

The goal of this paper is twofold. First, I investigate the existence of limits-to-arbitrage among U.S. ETFs investing in U.S. equity. Specifically, I examine the *extent* to which arbitrage costs affect mispricing both in the cross-section and in the time-series. Second, I analyze whether ETF returns comove excessively with other ETFs in similar styles. In this effort, I examine three style dimensions: size (large-, mid- and small-cap), value-growth and ETF liquidity (high-, medium- and low).

To preview my results, I find that, despite the existence of an *explicit* arbitrage mechanism, mispricing mean-reverts only *partially*; roughly by 25.2 %, 24.6 % and 21 % over a 1-day, 1-week, 4-week period respectively. A decomposition of this effect shows that mispricing stems mainly from the ETF, particularly at longer horizons suggesting that it is the ETF, rather than the NAV, that is the source of mispricing. I then analyze the extent to which proxies for arbitrage costs (transaction and holding costs) can explain the magnitude of mispricing. I find that mispricing *increases* with proxies for aggregate illiquidity, i.e. with the uncertainty in U.S. equity markets (VIX), with the volatility of the ETF's benchmark return and during the financial crisis (09/2008-03/2009). As an example, mispricing increased by 7 bps (or 24 % of the Std. Dev. of mispricing) for the average ETF during the crisis. Cross-sectional differences in ETF transaction costs seem to matter even more; the average difference in mispricing between funds in the top and bottom tercile of Amihud's illiquidity is 7.9 bps (30 % of Std. Dev.). Moreover, these cross-sectional differences are *amplified* during the financial crisis when the most illiquid ETFs experienced an increase in mispricing by 21 bps (72 % of Std. Dev.). Combined, the arbitrage cost proxies can explain more than 43 % of the variation in the magnitude of mispricing.

Mispricing is notoriously difficult to measure and it is not surprising that there is little direct evidence of its existence in prior studies, except for a handful of special cases such as closed-end country funds, dual-listed companies (or "Siamese twins"), index inclusions/exclusions and internet carve-outs and spin-offs (see the survey of Barberis and Thaler, 2003). Moreover, in all of these cases *fundamental risk* may still remain as arbitrageurs cannot fully hedge their portfolio due to a lack of perfect substitutes. My contribution is to remedy this deficiency in the literature by investigating ETFs, where mispricing is directly observable as the difference between the ETF market price and the fundamental NAV price, recorded synchronously at the end of each trading day. The U.S. ETF market can also be considered representative of the broader financial market based on: 1) its *size* ($1.2 trillion in total assets, out of which $632.5 billion in domestic U.S. equity in September 2012), 2) its share of U.S. equity dollar trading *volume* (roughly 30 % in 2012[2]) and disturbingly also because 3) ETFs have been shown to increase the *volatility* of the underlying assets and to *propagate* liquidity *shocks* to the underlying assets (Ben-David, Franzoni and Moussawi (2012); Shum *et al.* (2013)) .

To examine whether ETF returns comove *excessively* with each other, I begin by constructing excess ETF returns, defined as the difference between the ETF and the fundamental (NAV) returns. I then estimate a regression of excess ETF returns on the equally-weighted excess return of other ETFs with similar investment styles, and use the regression beta as my measure of excess comovement. The *information diffusion* view predicts excess comovements among the returns of ETFs with the highest liquidity differential vis-à-vis the underlying portfolio. Small-cap ETFs fit this description, while large-cap ETFs are not expected to have any significant liquidity improvement since the underlying securities are among the most liquid in the world. In contrast, the *correlated investor demand view* predicts excess comovement among ETF returns in similar investment styles. My results can differentiate between these two hypotheses; I find strong comovement patterns among *both* small- and large-cap ETFs, and limited comovement across categories. The economic magnitudes are also considerable. To illustrate, a 1 Std. Dev. increase in the equally-weighted excess return of own category ETF returns is, on average, associated with an increase in excess ETF returns by 8.5 bps (or 32 % of Std. Dev.) and 12.9 bps (or 59 % of Std. Dev.) for large- and small-cap ETFs respectively. After controlling for size-related excess comovement, I also find that the returns of value and growth focused ETFs covary excessively with each other. Similarly, ETFs with high and low liquidity seem to comove more strongly with other ETFs in their respective categories compared with the benchmark group of medium liquidity funds. This pattern also appears to be unrelated to either size, or value-growth related comovement. These results, along with the *stability* of excess comovements across return horizons, favor the correlated investor demand view of excess comovement.

[2] Financial Times on January 8th, 2013

In the existing literature, researchers are not in agreement on whether the asset pricing anomalies related to size or value-growth are the result of fundamental risk (e.g. the value premium is related to relative distress; Fama and French, 1993, 1994 and 1996) or unmodeled irrational behavior (as argued by Barberis and Thaler, 2003). Brav, Heaton, Li (2010) make an attempt to explain the factor premiums related to size and value-growth by arbitrage costs (as proxied by idiosyncratic volatility), but the results are mixed suggesting that limits to arbitrage may only provide a *partial* explanation for the survival of these factor premiums. Even if some of the common variation in size or value-growth stocks is related to fundamental risk, that is not a concern here as I have a direct way to control for variation in fundamentals. My contribution in this regard is to provide direct evidence of excess comovement in the returns of ETFs grouped by similar styles.

*Style investing* can lead to excess comovements in security returns. Barberis and Shelifer (2003) present a model where investors, to simplify decision making, first group assets into categories, and then allocate funds at the level of the category as opposed to at the individual asset level. If some of these category investors are also noise traders with correlated sentiment, and if their trading can impact prices, then as these investors move from one category to another based on past *relative* performance, their coordinated demand will induce a common factor in the returns of assets in the same category. The tendency to engage in category-based investing can result from an individual's need to categorize for the purpose of developing compact representations (or mental models) of the complex environment in which she is in (Mullainathan, 2002). An alternative interpretation of the style investing model is given by Barberis, Shleifer and Wurgler (2005), where investors have *preferred habitats* leading them to trade only a subset of all the available securities. When these investors' risk aversion, sentiment or liquidity needs change, they will adjust their exposure to the securities in their habitat, which will induce a common factor in the returns of these securities.

The strong demand for investment categories is evident from the large number of mutual funds, ETFs and hedge funds that follow *distinct* styles such as growth, value, and small-cap, and which are used by *both* individual and institutional investors (see e.g. Brown and Goetzmann (1997), Fung and Hsieh (1997), and Chan, Chen, and Lakonishok (2002)).

Cooper, Gulen and Rau (2005) find that mutual funds take advantage of current “hot” investment styles by renaming their fund to match the current hot style. These funds subsequently experience abnormal inflows with no improvement in performance. This hold even for cosmetic name changes, i.e. where the holdings of the fund do not change.

Consistent with the intuition that common variation in stock returns can be explained by correlated investor demand, there is growing empirical evidence suggesting that comovement in asset returns is related to the trading patterns of groups of investors. Kumar (2009) show how individual investors systematically shift their preferences across extreme style portfolios (small vs. large, value vs. growth) in response to past style returns, earnings differentials and advice from investment newsletters, while these shifts are found to be unaffected by innovations in macroeconomic fundamentals or changes in security specific expectations about cash-flows. Froot and Teo (2009) find evidence of style investing among institutional investors, whose reallocations across style groupings are found to be more intense than across random groupings of stocks. Moreover, they also find that own (distant) category inflows and returns predict future category returns positively (*negatively*) consistent with Barberis and Shleifer (2003). One important implication of too much short-run comovement in stock returns is that comovements must become *negative* at intermediate horizons so that, in the long run, prices reflect fundamentals. Consistent with this prediction, Greenwood (2008) constructs a simple trading strategy that bets on the reversion of the prices of over-weighted Nikkei 225 stocks that comove too much in the short-run and finds this trading strategy to yield significant risk-adjusted profits. Finally, Boyer (2011) find that economically meaningless value and growth index labels causes stock returns to covary excessively with each other following index reclassifications.

The article proceeds as follows. Section 2 gives a brief background on ETFs, their importance and the arbitrage mechanism. Section 3 describes the data and provides a definition of the key variables used in this study. Section 4 describes potential arbitrage costs in the ETF market and examines their importance for mispricing. Section 5 begins with a description of the key hypotheses and concludes with the main empirical findings. Section 6 provides additional robustness checks. Section 7 concludes.

## 2 Background on ETFs and Arbitrage

Exchange Traded Funds (ETFs) are investment vehicles that typically focus on replicating the performance of a pre-specified asset class, industry or geographical area. The first ever ETF, the Toronto Index Participation Unit (TIP) was created on March 9, 1990 in Toronto, though ETFs became popular only after the introduction of the S&P 500 ETF by State Street in 1993. This market has grown tremendously in the past 10 years with ETFs gathering more than $1.2 trillion in assets in the U.S., and $1.8 trillion globally, as of September 2012. This asset class is also capturing a large fraction of the transactions taking place in the financial markets. In the U.S., ETFs and other exchange traded products have been reported to account for roughly 30 % of all trading volume in 2012.

Before the proliferation of ETFs, most individual investors were limited to open- or closed-end funds and individual stocks. ETFs are hybrid instruments that combine the best features of both open- and closed-end funds. They combine the creation and redemption process of the former with the continuous tradability of the latter. ETFs have several other important advantages over open- and closed-end funds. Short-selling is often easier and cheaper than for the underlying assets due to the higher liquidity of ETFs, they generally have lower expense ratios than index funds and are considered tax efficient as the in-kind mechanism is a "non-taxable" event[3]. The main disadvantage of open-end funds is that they can be traded only once a day and they typically also suffer from a *cash drag* as the fund manager needs to keep some cash in hand for investor redemptions. In contrast, ETFs and closed-end funds can be traded throughout the day and both fund types avoid the cash-drag as investors can exit simply by selling the fund shares on the exchange at the prevailing market price. The main issue with closed-end funds is the lack of a creation or redemption mechanism – the number of shares is *fixed*, hence excess demand or supply may result in significant price deviations from the NAV. In fact, closed-end funds typically trade at a sizeable premium surrounding the IPO and move to a large discount within 6-months after

[3] In case of redemption, the ETF sponsor can minimize taxes by delivering the stocks with the largest embedded capital gains. Vanguard takes this one step further by embedding ETFs as a separate share class in an existing fund family. This allows the fund manager to select the securities with the largest embedded capital gains among *all* the funds in the same share class.

inception (Lee, Shleifer and Thaler, 1993). A unique feature of ETFs compared with other fund types is the "in-kind" share creation/redemption process, an *explicit* arbitrage mechanism that is designed to ensure efficient pricing.

## 2.1 The Arbitrage Mechanism

In a frictionless economy with rational investors, the market price of an ETF must equal its fundamental value, which is the market value of the underlying portfolio of securities held by the fund on a per-share basis (the *NAV*). The *level* of ETF mispricing is defined as the percentage deviation of the ETF price from the NAV:

$$PREM_{i,t} = \ln\left(ETF_t\right) - \ln\left(NAV_t\right) \tag{2.1}$$

For simplicity I call this measure a "premium" whether it is positive, or negative. The market price of an ETF may temporarily deviate from the NAV as a result of excess demand or supply. Such price deviations should not persist for long due to the in-kind creation/redemption process.

The in-kind process is only available to a select group of investors, typically large institutional investors, who have an agreement with the ETF sponsor. These investors, called Authorized Participants, are allowed to purchase or sell ETF shares in bundles (or "creation units") directly from the ETF sponsor in *exchange* for a basket of securities similar in composition to the benchmark index[4]. The creation process works as follows. During periods of strong *demand* for an ETF, its market price can be bid up in the market above the NAV. In order to take advantage of this price discrepancy, an Authorized Participant can purchase a portfolio of securities underlying the index, exchange this portfolio for ETF, sell the newly created ETF shares on the secondary market and realize an arbitrage profit that is equal to the difference between the ETF and NAV price, net of any transaction costs. Conversely, when there is strong *selling pressure* for an ETF, the ETF price can fall below the NAV. Authorized Participants can then purchase existing ETF

[4] For some ETFs redemptions occur in cash instead. These cases are mainly limited to some less liquid asset classes, such as international fixed income or commodities.

shares (at a discount), redeem them, and sell the securities received from the redemption to realize an arbitrage profit.

## 3 Data

My data selection starts with all U.S. traded Exchange-Traded Funds. I keep funds that 1) invest in domestic (U.S.) equity, 2) have underlying indices that are either equally- or value weighted[5], and 3) that have at least three years of data[6]. This data, along with NAVs and prices for the underlying benchmark, is obtained from Bloomberg. As reported by Petajisto (2011), Bloomberg coverage of ETFs is anywhere from 60 % to 97 % of all ETFs and 90 % to 99 % of all ETF assets. My second source is CRSP, which I use to obtain price and return data for all live and dead funds. The third data source is iShares covering daily NAV data for iShares funds from inception until 12/2012. iShares is the biggest provider of ETFs and it used to account for more than 50 % of all ETFs until 2005, and it generally still accounts for about 50 % of all assets. I prioritize NAV data from iShares over Bloomberg. Finally, I use Morningstar Direct to classify ETFs by market-cap and along the value-growth dimension. I end up with 159 U.S. equity ETFs; 21 small-, 49 mid- and 89 large-cap funds.

There are a handful of highly suspicious observations in the CRSP & Bloomberg data that I filter out. First, whenever ETF or NAV returns are more (less) than plus (minus) 20 % with a large reversal the next day, and no similar movement in the underlying benchmark index, I set the both the ETF and NAV returns equal to the return of the underlying index. Second, when mid-point ETF returns are more than 30 % greater than the ETF returns based on the closing price (in absolute value); I set the former equal to the latter. Third, when premiums are greater than 20 %, I set them equal to the previous day's value. Finally,

[5] This criterion is to ensure that the ETFs can be classified as "passive". Many recent ETFs track underlying indices that are either proprietary, or fundamentally weighted with the goal of achieving superior risk-adjusted returns. These ETFs blur the line between active and passive fund management.

[6] I exclude the first 6 months of a funds history since the data can often be unreliable or extreme. For instance, ETFs are highly illiquid when they are first created. Also, due to the low number of shares outstanding and the minimum fixed size of a creation/redemption basket, ETFs can experience dramatic creation/redemption activity early in the funds lifecycle. By excluding the first 6 months of a funds data, I essentially impose a 3.5 year requirement.

I winsorize all returns and premiums at seven standard deviations; estimated separately for each ETF.

The sample period starts on 31 May 2002 and ends on 31 December 2012. I chose this particular starting date to coincide with the date of a major index revision for all MSCI equity indices. These indices were adjusted to account for free float and the market coverage was extended from 60 % to 85 %. Tracking was more complicated before this index revision because the index was not fully investable as a large fraction of the outstanding shares are typically privately held and not accessible for trading. More importantly, the ETF market has undergone a dramatic expansion in scope and size in the last 10 years and earlier data may not be as representative of current market conditions.

## 3.1 Definitions of Key Variables

The main focus of this paper is on *excess* comovement in the returns of ETFs. The first step is to filter out the fundamental component of returns. One important advantage of using ETFs is that we don't need to estimate a Capital Asset Pricing Model. Instead, we can simply analyze "excess" ETF returns, which are defined as follows:

$$ETF_{i,t}^{e} = R_{i,t}^{ETF} - R_{i,t}^{NAV} \tag{2.2}$$

where the first (second) term is the return on the *ETF* (the *NAV*). ETF returns are calculated from the bid-ask midpoint (plus dividends) rather than using the returns given in CRSP, as suggested by Engle and Sarkar (2006). As it turns out, this measure of excess returns is equal to the *change* in mispricing, with the definition being exact on days when the ETF does not pay any dividends[7]. In order to determine whether comovement between ETF returns are excessive we need to establish that it is the ETF, rather than the NAV leg that is driving the results (see eq. 2.2). Hence we need to decompose the excess ETF return further.

[7] ETFs typically pay dividends every quarter or semi-annually.

Another dimension of ETF returns is related the extent to which the fund manager is able to *replicate* the performance of the underlying benchmark index. I define the replication error as follows:

$$RE_{i,t} = R_{i,t}^{NAV} - R_{i,t}^{UND} \tag{2.3}$$

where *NAV* is defined as before and *UND* refers to the underlying benchmark index. Replication errors may occur for a variety of reasons including; treatment of dividends, taxes, changes in index composition, securities lending and the purchase of only a subset of securities included in the underlying index. Changes in index composition due to additions or deletions, or because of supply and demand shocks (such as IPOs, SEOs, M&As) will force the ETF manager to trade in order to rebalance their portfolio. These amounts are not trivial; Petajisto (2011) documents that the median ETF generated an annual turnover of 29 % in 2010 by its *own* trading alone. Imperfect replication is also a concern for investors as a passive fund managers does not need to invest in all of the securities to replicate an index, but can instead use an optimized sampling technique to select a subset of securities that have the highest degree of comovement with the index.

Finally, investors are mainly concerned with the total *tracking deviation*, or the sum of the two dimensions of mispricing (eq. 2.2-2.3):

$$TD_{i,t} = ETF_{i,t}^{e} + RE_{i,t} = R_{i,t}^{ETF} - R_{i,t}^{UND} \tag{2.4}$$

This decomposition shows it is always possible to infer the third component from the other two. Moreover, if we analyze comovement patterns among two of the three components, it may be possible to identify the source of the comovement. To see this imagine a situation where there is a common factor in the returns of ETFs, but not in NAV returns. This will induce comovement between the excess ETF returns of two funds. Since the common factor is not present in NAV returns, the replication errors of any two ETFs will be uncorrelated with each other. A common component in excess ETF returns will then carry over to tracking deviations as per eq. (2.4). The bottom-line is that because both excess

ETF returns and tracking deviations include the ETF return, we can infer that the source of the excess comovement is the ETF, rather than the NAV.

## 3.2 Descirptive Statistics

Table 1 reports the descriptive statistics for the main variables used in this study. At the first year-end of the sample (12/2012), my sample of 55 ETFs accounted for almost $30 billion in Assets under management (AUM). By the final year-end (12/2012), there were 159 ETFs with more than $403 billion in AUM. ETF Liquidity, as measured by quoted spreads or Amihud's Illiquidity ratio[8] in the year 2012, was on average almost twice as high for small-cap funds as it was for large-cap funds, while mid-cap funds where comfortably in the middle. My liquidity measures are, however, positively skewed. Comparing medians reveals a different pattern; small- and large-cap funds now have similar spreads (4.9 and 42. bps), while mid-cap funds have the highest spreads (6.5 bps). Mid-cap funds are generally also smaller (in terms of AUM) than small-cap funds on a per-fund basis. Why is it that mid-cap funds are relatively illiquid and small? One reason could be that large-cap funds form the core portion of an investor's portfolio, while small-cap funds are more commonly used to capture the return premium related to size.

[Insert Table 1 about here]

Premiums are generally close to zero both at the mean and the median (see Panel B), but they do show a small, yet increasing pattern (in absolute terms) when moving from large-, to mid- and small-cap funds. However, mid-cap funds have the most volatile premiums (44 bps per day), while large- and small-cap funds are more comparable (26.2 and 27.9 bps). Similar patterns hold for Excess ETF returns. In contrast, replication errors are as volatile as excess ETF returns, but this seems to be driven by several extreme observations beyond the interquantile range. Some of this is probably explained by the fact that ETFs only pay dividends quarterly or semi-annually, while the underlying index pays them whenever the underlying constituents do. Index additions or deletions might also be expected to contribute due to their infrequent nature.

[8] See Appendix for definitions

[Insert Table 2 about here]

# 4 Limits to Arbitrage

In this section I begin by describing how various market frictions can limit the ability of arbitrageurs to take advantage of the explicit arbitrage mechanism (the in-kind creation/redemption process) that exists for ETFs. Without market frictions any price discrepancy should be eliminated instantaneously. Market frictions are therefore a *necessary* condition for the existence of excess comovements in the returns of ETFs. In the first empirical tests I show that mispricing stems mainly from the ETF, rather than the NAV. Finally, I empirically test the existence of limits-to-arbitrage by examining the relationship between arbitrage costs and the magnitude of mispricing both in the cross-section, and in the time-series.

## 4.1 Arbitrage Costs

In reality there are important obstacles due to market frictions that can limit the ability of arbitrageurs to profit from price discrepancies. Market frictions are often explicit and observable, such as transaction costs, holding costs and other trading restrictions (e.g. short-selling constraints). Both types of costs are expected to have similar relationship with mispricing; the greater the cost, the more severe is the mispricing.

*Transaction costs* are incurred every time a position is opened or closed. They include brokerage fees, commissions, creation/redemption fees, transactions taxes, bid-ask spreads and market impact costs. As described previously, the "in-kind" arbitrage mechanism involves the creation or redemption of ETF shares. A typical creation unit consists of 50 000 or 100 000 shares with dollar values ranging from $300 000 to $10 million. Every creation entails a fixed fee, usually $500 to $3000, which amounts to a few basis points of the creation value. Transaction costs for the underlying assets are generally very low for U.S. large-cap stocks, while they can be much higher for mid- and small-cap stocks. ETFs are often *more* liquid than their underlying portfolio of assets. To illustrate, consider the iShares Russel 2000 small-cap fund (TIC: IWM); the ETF is highly liquid with an average bid-ask spread of only 1 bp, but the underlying portfolio is relatively illiquid with an average spread of 10 basis points in September 2010[9]. Bid-ask spreads apply, however, only to limited quantities to be traded. Another dimension of trading costs is the market impact costs, i.e. larger transactions have a greater impact on transaction prices. This is also a concern in the ETF market where share creations/redemptions are typically very large: Petajisto (2011) reports that creations/redemptions occur on average once a week, while the average *size* of a creation/redemption transaction is about 21 % of AUM, or about 1556 % of average daily trading volume for the universe of U.S. traded ETFs during 2007-2010. As proxies for transaction costs, I use two measures, the monthly average of closing Quoted Spreads for normal sized transaction and Amihud's illiquidity ratio to capture the price impact dimension of transaction costs.

*Holding costs* refer to costs that accrue every period a position remains open. The most important one is idiosyncratic risk. Pontiff (2006) demonstrates theoretically that a rational investor's demand for a mispriced asset increases with the magnitude of mispricing, but decreases with the asset's idiosyncratic risk. His argument relies on the idea that the hedge position intended to capture the mispricing must, by design, be uncorrelated with the market and any other hedge positions. An arbitrageur will therefore be exposed to idiosyncratic risk whenever the arbitrageur has to *delay* liquidation the position.

In the context of ETFs, delays can occur because creations only take place at end-of-day NAVs, while the underlying portfolio may have to be purchased over an extended period of time to avoid large market impact costs (vice-versa for redemptions). Moreover, search and delay costs are more likely to arise when the position to be liquidated is large, as is often the case for creation/redemption transactions. For holding costs I use the idiosyncratic volatility of excess ETF returns similar to Gagnon and Karolyi (2010).

### 4.2 Source of Mispricing – ETF vs. NAV

The in-kind arbitrage mechanism ensures that the price of a mispriced ETF must *eventually* revert back to its fundamental value. Hence the premium of an ETF has to be mean-

[9] Blackrock - iShares Institutional Trading Report (September 2010)

reverting. This provides a simple testable prediction. Since the excess ETF return is (approximately) equal to the return difference between the ETF and the NAV, a positive *PREM* will by definition predict either a) a *smaller* cumulative ETF return, b) a *larger* cumulative NAV return or 3) both a smaller ETF return and a larger NAV return (i.e. a smaller excess ETF return). To investigate the third prediction, I estimate the following regression:

$$ETF_{i,t}^{e} = \alpha_{i,0} + \lambda_{i,1} PREM_{t-1} + e_{i,t} \tag{2.5}$$

where the dependent variable is the excess ETF return (see eq. 2.2) and *PREM* is the level of mispricing in period *t-1*. The null hypothesis is that $\lambda_{i,1} = 0$ and the alternative is $\lambda_{i,1} < 0$. If the alternative is true, it implies that a *positive* premium today predicts a *negative change* in the premium the next day (see the previous section on the equality between changes in premiums and excess ETF returns). I estimate the model using pooled OLS with ETF specific coefficients and with clustered standard errors by time.

[Insert Table 3 about here]

The regression results in Table 3 shows that excess ETF returns are indeed highly predictable and mean-reverting: at the daily level the *average* coefficient on *PREM* is -0.252 and it is significant at the 1 % level. Economically this means that roughly 25 % of a given level of mispricing reverses over a one-day period. Similar results hold at the weekly and monthly frequencies with the implication that shocks to mispricing occur relatively frequently. Interestingly, the results are strongest at the weekly horizon as indicated by both the higher rate of mean-reversion and adjusted R-squared, possibly because the signal-to-noise ratio is the highest at this frequency.

This result does not, by itself, tell us which leg of mispricing (ETF or NAV) is the source of predictability. To answer this question, I repeat the analysis with either the ETF or the NAV return as the dependent variable. The null hypothesis is the same for the ETF return as it is for excess ETF returns, while for NAV returns the null is that $\lambda_{i,1}=0$ and the alternative is $\lambda_{i,1}>0$.

There is some evidence to suggest that premiums predict ETF returns. While there is no meaningful relationship between the two variables for the average fund, premiums do have predictive power for the returns of small-cap ETFs at the daily (t-stat = 3.90) and monthly (t-stat = 3.76) horizons (see Table 2). Moreover, the coefficient estimates become increasingly negative with the *length* of the holding period consistent with the conjecture that ETF mispricing must *eventually* mean-revert and therefore predict ETF returns. Economically these effects are very strong. For instance, a one std. dev. increase in the premiums of small-cap ETFs is, on average, associated with a negative ETF return of 3.9 bp, 13.9 bp and 99.2 bps at the daily, weekly and monthly horizons respectively. Small-cap ETFs generally have the most illiquid underlying assets, and consequently also high arbitrage costs, which is why it not surprising to find the predictability to be concentrated here.

The results in Table 3 also suggest that the NAV component of returns is predictable at the daily horizon where the average $\lambda_{i,1}$ is statistically significant at the 5 % level. However, the effect is strongest for large-cap stocks and the statistical significance disappears for small-cap stocks. This suggests that predictability in NAV returns may stem from a source other than the mispricing of the NAV; Ben-David, Franzoni and Moussawi (2012) argue and find evidence consistent with ETFs being a catalyst for short-term investors, who are arguably more exposed to liquidity shocks than are the investors in the underlying securities. More importantly, they show that liquidity shocks from ETF prices are propagated to the underlying securities via the in-kind arbitrage mechanism. My finding that premiums predict NAV returns, particularly for large-cap ETFs that are generally highly liquid and therefore likely to attract short-term investors, is consistent with their hypothesis. This effect not only disappears in lower frequency data, but the effect is also reversed.

## 4.3 Arbitrage Costs and the Magnitude of Mispricing

In this section I test the hypothesis that mispricing is *higher* for funds with greater arbitrage costs. These costs tend to increase with the aggregate illiquidity in the financial markets for at least two reasons. First, low liquidity will by definition limit the profitability of ETF arbitrage via higher transaction costs. Second, low market liquidity may be an indication of low funding liquidity (as in Brunnermeier and Pedersen, 2009) suggesting that less capital is devoted to ETF arbitrage, which in turn leaves more room for mispricing to widen and persist. Since volatility and illiquidity are highly correlated, I use the standard deviation of the underlying ETF benchmark return (*VOLUND*) and the VIX index for the whole U.S. market to proxy for periods of high illiquidity. For funds that track the overall U.S. stock market, these two variables will be highly correlated in the time-series. I would therefore expect *VOLUND* to have greater explanatory power as it has both a time-series and cross-sectional dimension.

To better capture *cross-sectional* differences in arbitrage costs, I rank ETFs based on their average monthly liquidity (Quoted spread or Amihud's illiquidity ratio) or holding costs (idiosyncratic volatility) and assign funds into tercile portfolios. I then include the top (*HCOST*) and bottom (*LCOST*) tercile dummies in the regression. By omitting the second tercile dummy, the coefficients on *HCOST* and *LCOST* measure the relative effect w.r.t. this omitted group. Only one set of *HCOST-LCOST* dummies is included at a time since the different arbitrage cost dummies are relatively highly correlated. Finally, I also include fund specific intercepts to capture any omitted (and time-invariant) cross-sectional differences in mispricing. One such attribute is the size category of the benchmark ETF index – the constituent stocks in large-cap indices are generally much more liquid than those of small-cap indices.

[Insert Table 4 here]

To investigate the relation between mispricing and the aforementioned proxies for arbitrage costs I estimate the following regression:

$$\left|PREM_{i,t}\right| = \alpha_{i,0} + \lambda_{i,2}\left|PREM_{t-1}\right| + \lambda_3 VIX_t + \lambda_4 VOLUND_t + \lambda_4 HCOST_{t-1} + \lambda_5 LCOST_{t-1} + e_{i,t} \quad (2.6)$$

where the dependent variable is the *absolute* level of premium. I run the regression as a panel and cluster standard errors by the ETF. Table 4 presents the results when only the time-series variables are included. As expected, the magnitude of mispricing is positively related to both *VIX* and *VOLUND* at all frequencies (except *VIX* at the monthly horizon).

The effect of *VOLUND* on mispricing is relatively stable across horizons. For instance, a one Std. Dev. Increase in *VOLUND* is associated with an increase in the premium by 4.1 bps at the daily horizon (or 0.14 Std. Dev. of the premium). Turning the attention to the relation between the average premium and the liquidity of the underlying index (as measured by the size category), we can see that the premium increases from about 1.1 bps (4 % of Std. Dev.) for Large-Cap funds to 4 and 4.2 (9 % and 15 % of Std. Dev.) respectively for Mid- and Small-Cap funds at the daily level, all of which are highly statistically significant.

Next I estimate the full model with the cross-sectional dummies included. We can see a monotonic relationship between every proxy for arbitrage costs and the magnitude of mispricing (Table 5). For brevity I only report results for the daily frequency as the results are similar at both weekly and monthly investment horizons. The average difference in premiums between funds in the top and bottom tercile of illiquidity is 7.9 bps, or as much as 30 % of the standard deviation of the premium. Similar results hold when funds are ranked by Quoted Spreads, while the spread in premiums is only half as large between funds with high and low idiosyncratic volatility. This is not surprising since holding costs for ETFs arise only to the extent to which arbitrageurs transact in the underlying securities prior to the end-of-day creation event (or the following day in case of redemptions).

[Insert Table 5 here]

To provide further support to the argument that limits-to-arbitrage do exist in the market for ETFs, I investigate the behavior of premiums during the financial crisis when arbitrage costs were high across the board. Any cross-sectional differences should also be magnified if arbitrage costs increased proportionately more for illiquid vs. liquid ETFs, illiquid vs. liquid underlying indices (e.g. small-cap) or both. I extend the model in eq. (2.6) as follows:

$$\begin{aligned} \left|PREM_{i,t}\right| &= \alpha_{i,0} + \lambda_{i,2}\left|PREM_{t-1}\right| + \lambda_3 VIX_t + \lambda_4 VOLUND_t \\ &+ \lambda_{5,i} CRISIS + \left(\lambda_{60} + \lambda_{61} CRISIS\right) HCOST_{t-1} \\ &+ \left(\lambda_{70} + \lambda_{71} CRISIS\right) LCOST_{t-1} + e_{i,t} \end{aligned} \tag{2.7}$$

where *CRISIS* is an indicator variable for the financial crisis (9/2008-3/2009). First, I estimate a simplified version of the model by adding only the fund-specific *CRISIS* dummies to examine whether the average level of premiums changed during the crisis (Table 4: Panel B). The results support this hypothesis; premiums increased on average by 7 bps (or 0.24 Std. Dev. of PREM). Much of this increase is actually driven by mid-cap funds, which experienced an increase in mispricing by as much as 11.5 bps (or 28 % of Std. Dev.), while the mispricing of large- and small-cap funds increased by only 5.1 and 5.7 bps respectively. This large increase for mid-cap funds was most likely driven by the lack of liquidity in the ETF itself; during the crisis, the monthly average quoted spread for mid-cap funds was 63.9 bps compared with 39.6 bps and 41.2 bp for Large- and small-cap funds respectively.

The main hypothesis we are interested in is that funds with *high* arbitrage costs have higher absolute premiums during the financial crisis ($\lambda_{41} > 0$). To investigate this hypothesis, I estimate eq. (2.7) with the *CRISIS* dummy interacted with *HCOST* and *LCOST*. I now pool the coefficients for the *CRISIS* dummy together so that we can interpret the baseline effect of the omitted (medium cost) group. The results in Table 5 confirm the aforementioned prediction: the mispricing of ETFs in the top tercile of illiquidity *worsens* by about 21 bps (or 72 % Std. Dev.), while the premiums of ETFs in the middle and bottom terciles shows no consistent evidence of a decline. Similar results also hold when funds are ranked by quoted spreads, while differences in premiums between funds with high and low idiosyncratic volatility are only about half as large.

Overall, the results in this section confirm that the magnitude of mispricing is positively related to both time-series and cross-sectional determinants of arbitrage costs. Funds with the highest transaction and holding costs are much more prone to mispricing than other funds, especially during the financial crisis.

## 5 Excess Comovement in ETF Returns

Arbitrage costs determine the *boundaries* for mispricing. When the magnitude of mispricing is large enough to exceed this boundary, arbitrage becomes profitable and authorized participants can use the creation/redemption process to take advantage of this opportunity. So far I have shown how the boundaries for mispricing are wider for funds with high arbitrage costs, and during times when market volatility and illiquidity is high. However, as long as mispricing remains within the arbitrage costs boundaries, there is limited price pressure to correct for any price discrepancy. This may leave the ETF partially exposed to a common factor, causing the returns of ETFs to comove excessively with each other.

Comovement patterns in asset return have been reported among small-cap stocks, value/growth stocks, stocks within a geographical area and even among international markets. In frictionless markets with rational investors the ETF price will equal its fundamental value (the NAV), and any comovement in returns must be due to comovement in fundamentals. However, in economies with frictions or with irrational investors, and in which there are limits to arbitrage, comovement in returns may be partially delinked from fundamental giving rise to what is known as *excess comovement* (see Barberis and Shleifer, 2003; Barberis, Shleifer and Wurgler, 2005 and Greenwood, 2008). In the ETF market, the fundamentals of the ETF and the underlying portfolio are by definition the same. Hence we have a clear setting within which to examine the extent to which ETF returns comove excessively with each other and whether these patterns are related to arbitrage costs.

In the following section I lay out the theoretical predictions and explain the identifying assumptions. I then present the empirical strategy and provide the main results.

## 5.1 Key Hypotheses

There are two main theories that can explain why the returns of ETFs may comove excessively with each other due to a common factor. The first, the *information diffusion* view, states that, due to some market frictions, information is incorporated faster into the prices of ETFs as opposed to the underlying portfolio of securities (see e.g. Barberis, Shleifer and Wurger, 2005). This may occur because ETFs are generally less costly to trade than the underlying portfolio. According to this theory, there will be a common factor in the returns of securities that incorporate information at *similar* rates. This implies that the premium of an ETF reflects news that is incorporated into the price of the ETF, but that is not yet fully impounded into the prices of underlying securities. As an example, when good news about small-cap stocks is released towards the end of the trading day, an ETF that tracks a small-cap index will likely incorporate this information first, while the underlying portfolio will react with some delay, possibly overnight, resulting in a partially stale NAV and consequently also a positive premium. The liquidity improvement, or the difference in trading costs between the underlying portfolio and the ETF, is expected to be bigger for small-cap funds where the underlying stocks are less liquid. Hence, we would expect the impact of information dissemination to be most visible among small-cap ETFs.

**Hypothesis 3:** *The excess returns of small-cap ETFs comove with each other, but not with the returns of other ETFs.*

On the other hand, the underlying securities of large-cap U.S. ETFs are among the most liquid in the world. We would therefore *not* expect to see any material difference in the rate of information dissemination between large-cap ETFs and their underlying securities.

**Hypothesis 3b:** *There are no excess comovements between the returns of large-cap ETFs.*

The information diffusion view also predicts that return comovements should *decrease* with the length of the return horizon, simply because the informational advantage of ETFs relies on delays in information processing for the securities in the underlying portfolio. If the information in the ETF is incorporated into the underlying portfolio over a period of one or two trading days, as we would expect, then the strength of excess comovements should decline going from daily to weekly and monthly return horizons.

**Hypothesis 3c:** *The magnitude of excess comovements declines with the return horizon.*

The second set of theories builds on existence of *correlated investor demand or liquidity shocks* for a particular group of securities. Barberis and Shelifer (2003) present a model where investors, to simplify decision making, first group assets into categories (such as small-cap or growth stocks), and then allocate funds at the level of the category as opposed to at the individual asset level. If some of these category investors are also noise traders with correlated sentiment, and if their trading can impact prices, then as these investors move from one category to another, their coordinated demand will induce a common factor in the returns of assets in the same category.

**Hypothesis 4:** *The excess ETF returns of funds in similar categories comove with each other, but not with those in different categories.*

I use size and the value-growth dimension to identify similar styles as they are popular in the investment community. This claim is supported by the finding in Froot and Teo (2008) and Kumar (2009) that both retail and institutional investors allocate capital at the size and value-growth level.

In the model of Barberis and Shleifer (2003), correlated demand shocks are the result of noise trade sentiment. Sentiment is by definition mean-reverting, and hence excess comovements should *eventually* dissipate. These excess comovements need not, however, decline monotonically with the horizon over which returns are measured. For instance, if sentiment shocks are positively autocorrelated at the first few lags and negatively only thereafter, then excess comovements should *initially* increase with the return horizon and decrease only thereafter. Even if noise-trader sentiment is mean-reverting and excess comovements decline with the return horizon, this decline is likely to be *slower* than under the information diffusion view. This claim is supported by Barberis, Shleifer and Wurgler (2005); they find that comovements between newly added S&P 500 stocks increase with the remaining index stocks and this effect remains relatively *stable* across daily, weekly and monthly horizons, even after controlling for the effects of information diffusion, suggesting that noise-trader sentiment does not mean-revert quickly, but is instead rather persistent.

**Hypothesis 4b:** *The magnitude of excess comovements does not decline with the return horizon (up to 4 weeks).*

Barberis, Shleifer and Wurgler (2005) provide an alternative interpretation of the category view, whereby investors have *preferred habitats* leading them to trade only a subset of all the available securities. When these investors' risk aversion, sentiment or liquidity needs change, they will adjust their exposure to the securities in their habitat, which will induce a common factor in the returns of these securities. This model predicts that there will be a common factor in the returns of securities held and traded by a particular subset of investors, such as retail investors or investors with liquidity-related preferences. Ben-David, Franzoni and Moussawi (2012) have shown that ETFs are a catalyst for short-term investors. These investors' trades are at least partly motivated by liquidity needs and hence we would expect them to have a preference for highly liquid ETFs.

**Hypothesis 5:** *The excess ETF returns of highly liquid funds comove with each other.*

The liquidity of an ETF is dynamic and changes over time. Newly created ETFs are generally less liquid, while persistent inflows (outflows) will generally increase (decrease) the ETFs liquidity relative to other ETFs. For instance, when investors allocate funds based on the past relative performance, moving into styles that have performed relatively well in the past and out of styles that have performed relatively poorly (as in Barberis and Shleifer, 2003), the persistent in- and outflows will cause the relative liquidity of the ETFs in the former category to suffer at the expense of the latter. Hence, if we group ETFs by their liquidity, we will be implicitly grouping funds together that have experienced positive and negative inflows. In this dynamic setting we would not only expect that the most liquid funds comove excessively with each other, but also that the least liquid funds comove excessively with each other.

**Hypothesis 5b:** *The excess ETF returns will comove only with other ETFs with similar liquidity*.

Note that Hypothesis 5b encompasses hypothesis 5 and we can therefore differentiate between the two only if the former is found to be true. An alternative interpretation is simply that ETF liquidity is a style in itself (see e.g. Ibbotson *et al.* (2012)), in which case hypothesis 5b is a restatement of hypothesis 4. My main goal is to differentiate between the information diffusion view (Hypothesis 3) and the correlated investor demand or liquidity view (Hypothesis 4-5).

Finally I want to comment on an implicit assumption underlying the theories of correlated demand shocks when applied to the context of ETFs. Specifically, it is necessary to assume that there are *differences* in the clientele between the ETF and the underlying portfolio such that either clientele is *more* prone to category investing or habitat formation. If the same group of investors trade and hold both securities, then any demand or liquidity shock will affect both the ETF market price and the fundamental NAV price equally, resulting in no visible premium. This assumption of differential clienteles is likely to hold for several reasons. First, as documented by Ben-David, Franzoni and Moussawi (2012), domestic U.S. ETFs attract investors with shorter holding-periods in comparison to the underlying portfolio of securities. These investors are arguably more prone to correlated liquidity shocks. Second, retail investors will find it easier and cheaper to transact in the ETF market, rather than in the underlying securities, especially when dealing with less liquid securities. Thus we would expect there to be relatively more retail investors in the ETFs than in the underlying securities of these funds. Thirds, ETFs are often marketed as tools for investors who wish to gain a quick exposure to a particular market (see Economist 6/2011). Hence, it is easier for investors, particularly retail ones, to enter or switch categories by trading ETFs as opposed to constructing and trading such portfolios themselves. Finally, given the large variation in premiums over time, there is no reason to expect that correlation patterns among excess ETF returns exist unless there is a common factor in the returns of ETFs.

## 5.2 Results: Comovement within size based groups (Hypothesis 3 and 4)

I begin by investigating comovements between ETFs grouped by size, which is related to both the liquidity of the underlying index, as well as the degree of liquidity improvement between the ETF and the underlying index. As a measure of comovement, I use the regression beta from a regression of excess ETF returns on the equally-weighted excess returns of small-, mid- and large-cap ETFs ( $ETF^{e}_{k,t}$ , henceforth small-, mid- or large-cap factor):

$$ETF^{e}_{i,t} = \alpha_{i,0} + \lambda_{i,t-1} PREM_{i,t-1} + \beta_{i,k} \sum_{k=1}^{3} ETF^{e}_{k,t} + e_{i,t}, \qquad (2.8)$$

where k = 1 (Small), 2 (Mid) and 3 (Large) where the dependent variable is the excess ETF return (see eq. 2.2) and $PREM_{t-1}$ is the level of mispricing at the end of the previous period (see eq. 2.1). Since the factors are equally-weighted, the beta can be viewed as a measure of *average* comovement between the excess return of ETF *i* and the excess returns of those in group *k*. In order to avoid inducing a spurious correlation between ETF *i* and factor *k*, I compute the factor returns *excluding* ETF *i*. As in previous regressions, I estimate the model using pooled OLS with ETF specific coefficients and with clustered standard errors by time.

The *information diffusion view* says that the excess ETF returns of small-cap funds should comove with the small-cap factor, while the returns of large-cap funds should *not* comove with the large-cap factor. I test this hypothesis by examining whether the *average* betas are different from zero:

**Hypothesis 3 & 3b:** $\sum_{j=1}^{N_1} \beta_{j,1} / N_1 > 0$ and $\sum_{j=1}^{N_3} \beta_{j,3} / N_3 = 0$

where $N_k$ refers to the total number of funds in size category *k*. Since the model in eq. (2.9) is estimated using pooled OLS, I can simply use an F-test to determine whether the average beta is significantly different from zero.

The information diffusion view asserts that excess comovements arise because ETFs, particularly small-cap focused, incorporate information faster than the underlying securities. The information diffusion view implies that the premium of an ETF reflects news that is incorporated into the price of the ETF, but that is not yet fully impounded into the prices of underlying securities. This implies that premiums should predict NAV returns positively, particularly for small-cap funds that have relatively illiquid underlying portfolios. As we already saw in Table 3, this prediction is not borne out in the data.

The *category based view* asserts that the excess returns of funds in similar styles (e.g. based on size) comove with each other, but not with the returns of funds in different styles:

**Hypothesis 4:** $\sum_{k=1}^{N_k} \beta_{j,k} / N_k > 0$ and $\sum_{j \notin k} \beta_{j,k} / \left(N - N_k\right) = 0$

where $N$ is the total number of funds in the sample, and $N$ - $N_k$ is the number of funds not in category *k*. I focus mainly on the large- and small-cap categories as they are natural opposites to another.

[Insert Table 6 about here]

Table 6 provides the coefficient estimates averaged across all funds (Panel A), and averaged category-by-category (Panel B). The results in Panel A show that the average beta on the large-cap factor is highly statistically significant and economically strong – at the weekly level the average large-cap beta is almost 3.3 and 2.4 times as large as the average beta for Mid- and Small-Cap after accounting for differences in variability across the three factors[10]. Even if these *differences* in average betas are driven by the unequal number of funds in each size category, this evidence is nevertheless consistent with the category based view as it is the *only* one that predicts a significant large-cap factor loading in the first place. Furthermore, these comovement patterns persist over weekly and monthly horizons and the adjusted $R^2$s increase by almost half for every horizon in comparison to the base model that only includes the lagged premium (see Table 3). As an example, controlling for the three size factors increases the explanatory power of the regression from 24.40 % to 36.24 % at the weekly horizon.

The category-by-category results (Panel B) provide more direct evidence in support of the category based view: the excess returns of large-cap funds comoves *only* with the large-cap factor (F-stat = 187.63), and small-cap funds comove only with their own factor (F-stat = 284.67). The economic magnitudes of the coefficient estimates are also considerable. To illustrate, a 1 Std. Dev. increase in the *own* category factor is on average associated with an increase in excess ETF returns by 8.5 bps (or 0.32 Std. Dev.) and 12.9 bps (or 0.59 Std. Dev.) for large- and small-cap ETFs respectively. These numbers are for the weekly return horizon, but the results remain similar at daily, and to a lesser extent even at monthly return horizons. This stability in results over longer return horizons is consistent with my conjecture that correlated demand shocks caused by noise trader sentiment are quite persistent (Hypothesis 4b).

While Mid-cap ETFs comoves strongly with their own factor, they also comove to some extent with the large- and small-cap factors. This effect is mainly visible in daily data,

[10] More specifically, I scale down the large-cap factor by: Std. Dev (large cap factor)/Std. Dev(Mid- or Small-Cap factor).

while the own-category effect becomes stronger at weekly and monthly horizons. Mid-cap ETFs generally have lower liquidity than either large- or small-cap ETFs. Hence we would not expect there to be any meaningful improvement in liquidity over the underlying portfolio, and also no possibility for the information diffusion hypothesis to hold in this case. This finding of excess comovements among the returns of mid-cap ETFs, albeit is weak, is nevertheless in support of the category based view.

To ensure that my finding of excess comovements between small-, mid- and large-cap funds is not driven by differences in the liquidity of the ETF, I include interaction terms for low and high illiquidity ETFs with each size-based factor loading in eq. (2.9). I find some evidence to suggest that more liquid ETFs comove more strongly with the large-cap factor and vice-versa for less liquid ETFs, but this effect is not robust across holding periods or when quoted spreads are used instead of Amihud's *ILLIQ*. Even when the results do hold, the economic magnitude of the own-category based comovements barely changes.

Finally, I also make an attempt to verify that the excess comovements documented here are driven by the ETF, rather than the NAV component of returns, I estimate eq. (2.9) with replication errors as the dependent variable[11]. The average factor betas are generally insignificant in weekly and monthly data. In daily data the betas are oftentimes negative and significant, particularly for small-cap ETFs, but the effects are concentrated in the distant-categories, while the own-category comomvenents are barely affected.

## 5.3 Results: Comovement within style based groups (Hypothesis 4)

In this section I examine whether there are excess comovements among value and growth focused ETFs. The liquidity of the *underlying* index is likely to be similar across these two styles since there is roughly an equal number of large-, mid- and small-cap funds among both value and growth ETFs (see Table 1). While there may be some differences in the average ETF *liquidity* of value vs. growth funds, this is to be expected if there are category investors who move in and out of styles based on their relative past performance, inducing

[11] Results available upon request

in- or outflows from one category into another (as described in Barberis and Shleifer, 2003).

I examine the extent of excess comovement among value and growth styles in the context of a regression of excess ETF returns on the equally-weighted excess returns of value, growth and blend ETFs:

$$ETF^{e}_{i,t} = \alpha_{i,0} + \lambda_{i,t-1} PREM_{i,t-1} + \beta_{i,k} \sum_{k=1}^{3} ETF^{e}_{k,t} + e_{i,t}, \tag{2.9}$$

where k = 1 (Blend), 2 (Value) and 3 (Growth)

The *category based view* asserts that value and growth ETFs comove within their own respective styles:

**Hypothesis 4:** $\sum_{k=1}^{N_2} \beta_{j,2} / N_2 > 0$ and $\sum_{k=1}^{N_3} \beta_{j,3} / N_3 > 0$

The blend category is by definition neither value, nor growth, and therefore it is not a style. But there are a disproportionate number of large-cap funds in the blend category. It would therefore not be surprising to find some evidence of excess comovement within the blend category as well. In fact, each of the three categories blend, value and style have a disproportionate fraction of large-cap funds. We would therefore also expect to see some degree of comovement between value and blend, and growth and blend. Viewed in this way, the blend factor is simply a size factor in disguise, tilted towards large-cap funds. It is important to control for this factor to avoid miss-interpreting any style related comovement as being related to size.

[Insert Table 7 about here]

Table 7 provides the coefficient estimates averaged across all funds (Panel A), and averaged style-by-style (Panel B). If size is the only style that matters, then we would expect only the beta on the blend factor to be significant. The average beta on the blend factor is indeed highly significant, however, so are both the Value and Growth betas at the daily and weekly horizons. At the monthly level the value factor loses its significance. This finding masks a key category-specific effect: value ETFs comove significantly with other value ETFs (F-stat = 57.29), while Growth and Blend ETFs comove *negatively* (insignificant) with the value category. The main finding that holds across all horizons is that both value and growth ETFs commove significantly with their own respective styles, while the comovement with the opposite style is weak at best. Economically a 1 Std. Dev. shock to the Value factor returns is associated with an increase in excess returns by 7.2 bps (or 0.23 Std. Dev) for funds in the same style and at the weekly horizon, while for growth funds the increase in excess returns is about 4.7 bps (or 0.12 Std. Dev.). I also conducted an F-test for the equality of means between the average Value and Growth betas. For Value ETFs the average betas on the Value and Growth factors are statistically different from each other, while for Growth ETFs this is only true at the monthly horizon.

I also verify that the results are not driven by the NAV component of returns by estimating eq. (2.11) with replication errors as the dependent variable[12]. There is some evidence to suggest that replication errors for all categories are positively related to the growth factor. Consequently, I might be underestimating the excess comovement within growth ETFs, and overestimating it for value ETFs.

In summary, the excess ETF returns of value and growth funds comove positively with other ETFs in the same style, even after accounting for size-related comovement via the blend factor that has a similar proportion of large-, mid- and small-cap funds as do the value and growth ETFs. Moreover, value and growth funds comove weakly at best with ETFs in the opposite style. This finding supports the category based view whereby investors simplify decision making by first categorizing securities into styles, and then allocate funds at the category level. Value and growth focused ETFs offer investors an easy and cheap way to achieve this goal. This could be the reason why these ETFs appear to be *more* prone to excess comovement *in comparison* to the underlying benchmark portfolios.

## 5.4 Results: Comovement within illiquidity based groups (Hypothesis 5)

ETFs are a catalyst for short-term investors who are arguably more exposed to liquidity shocks than are the investors of the underlying securities. In this section I investigate the

[12] results available upon request

hypothesis that the short-term nature of these ETF investors trades' induce them to hold liquid ETFs (Hypothesis 5), or in a dynamic setting to trade ETFs that have experienced shifts in liquidity from high to low (Hypothesis 5b).

I begin by grouping ETFs into low-, medium- and highly illiquid funds based on a ranking of monthly average illiquidity (Quoted spread or Amihud's illiquidity ratio). The identity of funds in each category may *change* over time as funds experience relative increases or declines in illiquidity. I estimate a regression of excess ETF returns on the equally-weighted excess returns of ETFs with low-, medium- and high illiquidity (henceforth low, medium and high *ILLIQ* factor) :

$$ETF_{i,t}^{e} = \alpha_{i,0} + \lambda_{i,t-1} PREM_{i,t-1} + \left(\beta_{i,k} + \delta_{L,1} I_{L=1} + \delta_{L,3} I_{L=3}\right) \sum_{L=1}^{3} ETF_{L,t}^{e} + e_{i,t}, \qquad (2.10)$$

where L = 1 (Low), 2 (Medium) and 3 (High ILLIQ)

where the $ETF_{L,t}^{e}$ is the equally-weighted excess return of funds ranked by either Amihud's illiquidity ratio or Quoted Spreads. Since the identity of funds in the three categories changes over time I cannot simply average the betas category-by-category as I did previously. Instead I interact each *ILLIQ* factor with two dummy variables; the first takes the value 1 whenever ETF *i* belongs to the illiquidity category 1 (*Low*) and the second takes value whenever ETF *i* belongs to illiquidity category 3 (*High*). By omitting the second category, the coefficients $\delta_{L,1}$ and $\delta_{L,3}$ measure the degree of comovement with factor *L* *relative* to the omitted group.

In a static setting, hypothesis 5 predicts that funds with the *highest* liquidity comove with other highly liquid funds ($\beta_{11} > 0$), while the dynamic prediction of hypothesis 5b states that the returns of *both* highly liquid and illiquid funds comove excessively within their own category ($\beta_{11} > 0$, $\beta_{13} = 0$ and $\beta_{33} > 0$, $\beta_{31} = 0$).

[Insert Table 8 about here]

Table 8 provides the coefficient estimates for $\beta_{i,k}$ averaged across all funds (Panel A). If there is comovement only between funds in the same liquidity category (as predicted by hypothesis 5b), and if the variability in the factors is similar (which it is), then we would expect the average betas for the three categories to be roughly of the *same* magnitude. This is indeed what the results show: the impact of a 1 Std. Dev. increase in low, medium and high *ILLIQ* factors is associated with an increase in excess ETF returns by 3.2, 2.8 and 2.8 bps (9.4 %, 8.1 %, 8.2 % of Std. Dev.) for weekly data. The average factor betas are also all significant at the 1 % level.

Turning the attention to the category-specific dummies (Panel B), the results paint a clear picture of significant comovement among ETFs in the low and high ILLIQ categories. To illustrate, consider the liquidity factor based on Quoted Spreads: low illiquidity funds comove significantly more with the low *ILLIQ* factor (t-stat = 5.58) and significantly less with the high *ILLIQ* factor (t-stat = -3.03) when compared with the *omitted* medium illiquidity group. Similarly, highly illiquid ETFs comove significantly more with the high *ILLIQ* factor (t-stat = 2.86) and less with low *ILLIQ* factor (t-stat = -4.49). Finally, *both* high and low illiquidity ETFs comove significantly *less* with the medium *ILLIQ* factor compared with ETFs in this category.

In further robustness test I also verify that the ILLIQ factors do not simply mimic the three size factors that I used in the previous section. I do this by investigating whether there are any systematic differences between large-, mid- and small-cap ETFs in terms of their betas on the three ILLIQ factors. The results do not show any distinct patterns[13].

Finally, I also verify that the results are not driven by the NAV component of returns by estimating eq. (2.10) with replication errors as the dependent variable[14]. The average factor betas are mainly insignificant across all holding periods and measures of liquidity. In terms of the category-dummy interactions with the factors, there are no consistent results across horizons or measures of liquidity suggesting that the excess comovements among liquidity groupings is not driven by replication errors.

[13] Available upon request
[14] results available upon request

## 6 Summary and Conclusions

In this paper I have analyzed two main questions: First, to what *extent* do limits-to-arbitrage explain the mispricing of U.S. based Exchange-Traded Funds investing in U.S. equity? In the context of ETFs, mispricing is directly observable as the difference between the ETF market price and its fundamental (NAV) price. Second, is there a *common factor* in the returns of ETFs related to size, value-growth or ETF liquidity, but unrelated to fundamentals? In other words, do ETF returns comove *excessively* with the returns of other ETFs in similar styles after controlling for the fundamentals (the NAV returns)?

ETFs are unique in the sense that there is an *explicit* arbitrage mechanism (the in-kind creation/redemption process) that should, in frictionless markets, ensure pricing efficiency. Despite the existence of this mechanism, I find that mispricing mean-reverts only *partially* and this effect weakens with the length of the return horizon. A decomposition of this effect shows that mispricing stems mainly from the ETF, particularly at longer horizons suggesting that it is the ETF, rather than the NAV that is the source of mispricing. I then analyze the extent to which proxies for arbitrage costs (transaction and holding costs) can explain the magnitude of mispricing. In the time-series, I find that mispricing increases with proxies for aggregate *illiquidity* (i.e. VIX, Std. Dev. of the ETFs benchmark return and during the financial crisis in 09/2008-03/2009), but cross-sectional differences in ETF transaction costs seem to matter even more. These two dimensions are also found to *amplify* each other. Combined, the arbitrage cost proxies can explain more than 43 % of the variation in the magnitude of mispricing.

To examine whether ETF returns comove *excessively* with each other, I estimate a regression of excess ETF returns on the equally-weighted excess return of other ETFs with similar investment styles, and use the regression beta as my measure of excess comovement. The *information diffusion* view predicts excess comovements among the returns of ETFs where the liquidity differential with the underlying portfolio is the greatest. Small-cap ETFs fit this description, while large-cap ETFs do not. In contrast, the *correlated investor demand view* predicts excess comovement among ETF returns in similar investment styles. My results can differentiate between these two hypotheses; I find strong comovement patterns among *both* small- and large-cap ETFs, and limited comovement across categories. After controlling for size-related excess comovement, I also find that the returns of value and growth focused ETFs covary excessively with each other. ETF liquidity appears to be another distinct source of commonality consistent with prediction that ETFs attract short-term investor with a preference for liquidity. These results remain stable over daily, weekly and monthly return horizons.

Overall, the results in this paper highlight that limits-to-arbitrage are important predictors of mispricing even though ETFs have an explicit arbitrage mechanism that is designed to ensure pricing efficiency. My results also paint a clear picture of excess comovement among ETF returns in similar investment styles consistent with recent theoretical work on the consequences of commonality in trading behaviour by a group of investors.

## References


Ackert, L.F. and Tian, Y.S. (2008): Arbitrage, Liquidity, and the Valuation of Exchange Traded Funds, *Financial markets, institutions & instruments*, 17 (5), pp. 331-362.

Barberis, N., & Shleifer, A. (2003). Style investing. *Journal of Financial Economics*, 68(2), 161-199.

Barberis, N., Shleifer, A., & Wurgler, J. (2005). Comovement. *Journal of Financial Economics*, 75(2), 283-317.

Barberis, N., & Thaler, R. (2003). A survey of behavioral finance. *Handbook of the Economics of Finance*, 1, 1053-1128.

Ben-David, I., Franzoni, F., and Moussawi, R. (2012): *ETFs, Arbitrage, and Shock Propagation*. Fisher College of Business Working Paper No. 2011-03-20; Charles A. Dice Center Working Paper No. 2011-20; Swiss Finance Institute Research Paper No. 11-66; AFA 2013 San Diego Meetings Paper. Available at SSRN: http://ssrn.com/abstract=1967599 or http://dx.doi.org/10.2139/ssrn.1967599

Boyer, B. H. (2011). Style‐Related Comovement: Fundamentals or Labels?. *The Journal of Finance*, 66(1), 307-332.

Bodurtha, J.N., Kim, D-S., Lee, C.M.C (1995): Closed-end Country Funds and U.S. Market Sentiment, *The Review of Financial Studies*, Vol. 8, No. 3, pp. 879-918.

Brav, A., Heaton, J. B., & Li, S. (2010). The limits of the limits of arbitrage. *Review of Finance*, 14(1), 157-187.

Brown, S. J., & Goetzmann, W. N. (1997). Mutual fund styles. *Journal of financial Economics*, 43(3), 373-399.

Brunnermeier, M., Pedersen, L. (2009): Market liquidity and funding liquidity. *Review of Financial Studies*, 22, pp. 2201–2238.

Chan, L. K., Chen, H. L., & Lakonishok, J. (2002). On mutual fund investment styles. *Review of financial studies*, 15(5), 1407-1437.

Cooper, M. J., Gulen, H., & Rau, P. R. (2005). Changing names with style: Mutual fund name changes and their effects on fund flows. *The Journal of Finance*, 60(6), 2825-2858.

DeLong, J. B., Shleifer, A., Summers, L.H. and Waldmann, R. (1990): Noise trader risk in financial markets. *Journal of Political Economy*, 98(4):703-738.

Engle, R. and Sarkar, D. (2006): Premiums-Discounts and Exchange Traded Funds. *The Journal of Derivatives*, 2006 (1), pp. 27-45.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F., and Kenneth R. French, 1995, Size and book-to-market factors in earnings and returns, *Journal of Finance,* 50, 131–155.

Fama, Eugene F., and Kenneth R. French, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance,* 51, 55–84.

Froot, K., & Teo, M. (2008). Style investing and institutional investors. *Journal of Financial and Quantitative Analysis*, 43(4), 883.

Greenwood, Robin. (2008): "Excess comovement of stock returns: Evidence from cross-sectional variation in Nikkei 225 weights." *Review of Financial Studies,* 21.3, pp. 1153-1186

Hardouvelis, G. La Porta, R. Wizman, T. A. (1994): *What Moves the Discount on Country Equity Funds*? J.A. Frankel (Ed.), The Internationalization of Equity Markets (pp. 345-403), Chicago, IL: University of Chicago Press.

Ibbotson, Roger G., Chen, Zhiwu, Kim, Daniel Y.-J. and Hu, Wendy Yunchun (2012): Liquidity as an Investment Style. Available at SSRN: http://ssrn.com/abstract=1817889.

Kumar, A. (2009). Dynamic style preferences of individual investors and stock returns. *Journal of Financial and Quantitative Analysis*, 44(03), 607-640.

Lee, C.M., Shleifer, A. and Thaler, R.H. (1991): Investor Sentiment and the Closed-End Fund Puzzle. *The Journal of Finance*, XLVI (1), pp. 75-109.

Mullainathan, S. (2002). *Thinking through categories*. Unpublished working paper, Massachusetts Institute of Technology.

Baruch, S., Andrew Karolyi, G., & Lemmon, M. L. (2007). Multimarket trading and liquidity: theory and evidence. *The Journal of Finance*, 62(5), 2169-2200.

Petajisto, A (2011): *Inefficiencies in the Pricing of Exchange-Traded Funds*. Available at SSRN: http://ssrn.com/abstract=2000336 or http://dx.doi.org/10.2139/ssrn.1572907

Pontiff, J. (2006). Costly arbitrage and the myth of idiosyncratic risk. *Journal of Accounting and Economics*, 42(1), 35-52.

Shleifer A., and Vishny, R. (1997): The limits to arbitrage. *Journal of Finance*, 64, pp. 1517-1548.

Rodier, A., Haryanto, E., Shum, P., & Hejazi, W. (2013). *Intraday Share Price Volatility and Leveraged ETF Rebalancing*. Available at SSRN: http://ssrn.com/abstract=2161057.

**Table 1: Snapshot of ETF statistics in 2012**

| | Nr. of ETFs | AUM (in million $) | QSPR (bp) | | ILLIQ | | ID. Vol. | |
|---|---|---|---|---|---|---|---|---|
| | | | AVG | MED | AVG | MED | AVG | MED |
| **Large Cap** | **89** | **266,658** | **6.1** | **4.9** | **2.3** | **0.22** | **4.2** | **3.4** |
| Large Value | 22 | 60,874 | 6.0 | 4.9 | 2.6 | 0.14 | 4.2 | 3.6 |
| Large Blend | 38 | 116,691 | 5.8 | 4.7 | 3.0 | 0.22 | 4.0 | 3.4 |
| Large Growth | 29 | 89,092 | 6.7 | 5.3 | 1.1 | 0.29 | 4.5 | 3.3 |
| **Mid Cap** | **48** | **85,212** | **9.0** | **6.5** | **8.0** | **0.37** | **6.2** | **4.0** |
| Mid Value | 14 | 10,868 | 10.1 | 7.7 | 13.3 | 0.95 | 4.4 | 3.7 |
| Mid Blend | 18 | 57,653 | 10.1 | 5.0 | 9.8 | 0.19 | 6.1 | 4.2 |
| Mid Growth | 22 | 16,691 | 6.8 | 6.2 | 1.2 | 0.34 | 8.0 | 3.8 |
| **Small Cap** | **22** | **51,590** | **11.5** | **4.2** | **4.2** | **0.12** | **7.1** | **5.0** |
| Small Value | 6 | 9,320 | 17.5 | 3.9 | 5.8 | 0.04 | 10.7 | 5.4 |
| Small Blend | 9 | 31,720 | 12.3 | 5.8 | 5.5 | 0.27 | 6.5 | 5.3 |
| Small Growth | 7 | 10,550 | 5.2 | 3.9 | 1.1 | 0.12 | 4.6 | 4.2 |
| All | 159 | 403,459 | 7.7 | 5.2 | 4.3 | 0.20 | 5.2 | 3.8 |

*Note:* AVG and MED refer to the average or median of the monthly Quoted spread/Amihud's Illiquidity ratio/ Idiosyncratic volatility during year 2012. AUM refers to the monthly average of assets under management during December 2012.

**Table 2: Descriptive Statistics**

| Variable - Category | Mean | P25 | MED | P75 | Std. Dev. |
|---|---|---|---|---|---|
| **Premium** | **-0.002** | **-0.073** | **0.000** | **0.070** | **0.321** |
| Large-Cap | 0.002 | -0.064 | 0.000 | 0.068 | 0.262 |
| Mid-Cap | -0.001 | -0.079 | -0.000 | 0.073 | 0.440 |
| Small-Cap | -0.016 | -0.102 | -0.010 | 0.076 | 0.279 |
| **Excess ETF Return** | **0.000** | **-0.044** | **0.000** | **0.045** | **0.265** |
| Large-Cap | 0.000 | -0.039 | 0.000 | 0.039 | 0.185 |
| Mid-Cap | 0.000 | -0.048 | 0.000 | 0.048 | 0.410 |
| Small-Cap | 0.000 | -0.067 | 0.000 | 0.068 | 0.190 |
| **Replication Error** | **0.002** | **-0.006** | **-0.000** | **0.007** | **0.223** |
| Large-Cap | 0.002 | -0.005 | -0.000 | 0.007 | 0.136 |
| Mid-Cap | 0.001 | -0.006 | -0.001 | 0.008 | 0.356 |
| Small-Cap | 0.001 | -0.006 | 0.000 | 0.008 | 0.185 |

**Table 3: Mean-reversion**

| | *Excess ETF returns* | | |
|---|---|---|---|
| | **Daily** | **Weekly** | **Monthly** |
| Mean-reversion | -0.252 | -0.246 | -0.210 |
| | (171.55)*** | (73.68)*** | (32.82)*** |
| Large-Cap | -0.246 | -0.249 | -0.202 |
| | (144.65)*** | (57.91)*** | (35.91)*** |
| Mid-Cap | -0.259 | -0.216 | -0.212 |
| | (154.12)*** | (72.50)*** | (42.30)*** |
| Small-Cap | -0.261 | -0.298 | -0.240 |
| | (160.50)*** | (49.65)*** | (8.41)*** |
| Adj. R2 | 0.234 | 0.240 | 0.152 |
| Obs | 309,685 | 63,998 | 16,055 |

| | *ETF Returns* | | | *NAV returns* | | |
|---|---|---|---|---|---|---|
| | **Daily** | **Weekly** | **Monthly** | **Daily** | **Weekly** | **Monthly** |
| Mean-reversion | 0.012 | -0.487 | -2.706 | 0.328 | -0.186 | -2.404 |
| | (0.01) | (1.80) | (1.27) | (6.59)** | (0.26) | (0.99) |
| Large-Cap | 0.114 | -0.276 | -2.046 | 0.425 | 0.040 | -1.755 |
| | (0.64) | (0.55) | (0.81) | (7.72)** | (0.01) | (0.59) |
| Mid-Cap | -0.056 | -0.796 | -2.775 | 0.266 | -0.540 | -2.468 |
| | (0.20) | (3.02)* | (0.97) | (4.02)* | (1.37) | (0.76) |
| Small-Cap | -0.248 | -0.667 | -5.227 | 0.070 | -0.327 | -4.886 |
| | (3.90)* | (2.24) | (3.76)** | ( 0.32) | (0.53) | (3.22)* |
| Adj. R2 | 0.002 | 0.001 | 0.015 | 0.007 | 0.002 | 0.015 |
| Obs | 309,685 | 63,998 | 16,055 | 309,685 | 63,995 | 16,055 |

*Note:* This table reports the *average* mean-reversion coefficient and the associated F-statistic in parenthesis. Standard errors are corrected for clustering by time. */**/*** indicates statistical significance at the 10, 5 and 1 % level.

**Table 4: Explaining Mispricing**

| Variables | DAILY | WEEKLY | MONTHLY |
|---|---|---|---|
| ***Panel A: Average Mispricing*** | | | |
| VIX | 0.0036 | 0.0021 | 0.0019 |
| | (15.54)*** | (4.20)*** | (1.27) |
| Std. Dev. of UND | 0.0370 | 0.0406 | 0.0449 |
| | (6.89)*** | (6.37)*** | (3.34)*** |
| Large Cap | 0.011 | 0.017 | 0.017 |
| | (5.14)** | (0.57) | (0.57) |
| Mid Cap | 0.040 | 0.037 | 0.041 |
| | (66.03)*** | (11.72)*** | (3.11)* |
| Small Cap | 0.042 | 0.037 | 0.047 |
| | (79.99)*** | (14.71)*** | (4.98)** |
| Adjusted R2 | 0.435 | 0.418 | 0.399 |
| *N* | 309,685 | 63,372 | 15,801 |
| ***Panel B: Average Mispricing During the Financial Crisis*** | | | |
| *CRISIS* | 0.070 | 0.023 | 0.029 |
| | (41.28)*** | ( 5.73)** | (1.61) |
| Large Cap (*CRISIS)* | 0.051 | 0.039 | 0.029 |
| | (19.83)*** | (3.23)* | (0.31) |
| Mid Cap (*CRISIS)* | 0.115 | 0.122 | 0.114 |
| | (63.24)*** | (11.28)** | (2.22) |
| Small Cap (*CRISIS)* | 0.057 | 0.046* | 0.050 |
| | (17.65)*** | (2.88) | (0.44) |
| Adjusted R2 | 0.344 | 0.337 | 0.305 |
| N | 297,167 | 60,803 | 15,139 |

*Note:* The number in parenthesis is the F-statistics ($H_0$: *average* coefficient = 0), except for VIX and Std. Dev. of UND. The *average* intercept for large-, mid- and small-cap funds is reported in Panel A. *CRISIS* refers to the *average* coefficient on the CRISIS dummy, while Cap (CRISIS) is the average coefficient for each size category. Standard errors are corrected for clustering by time. */**/*** indicates statistical significance at the 10, 5 and 1 % level.

**Table 5: Mispricing and cross-sectional ETF determinants**

| Variables | QSPR | | ILLIQ | | Idiosyncratic Volatility | |
|---|---|---|---|---|---|---|
| LCOST | -0.0234 | -0.0227 | -0.0261 | -0.0273 | -0.0118 | -0.0120 |
| | (20.14)*** | (18.44)*** | (17.56)*** | (17.98)*** | (12.67)*** | (13.54)*** |
| HCOST | 0.0384 | 0.0265 | 0.0531 | 0.0399 | 0.0149 | 0.0097 |
| | (24.61)*** | (15.49)*** | (26.93)*** | (19.62)*** | (11.86)*** | (7.51)*** |
| CRISIS | | -0.0044 | | -0.0299 | | 0.0330 |
| | | (0.38) | | (2.71)*** | | (2.75)*** |
| LCOST * | | -0.0009 | | 0.0280 | | -0.0035 |
| CRISIS | | (0.16) | | (4.09)*** | | (0.53) |
| HCOST * | | 0.1910 | | 0.2377 | | 0.0784 |
| CRISIS | | (12.21)*** | | (12.35)*** | | (8.06)*** |
| Adjusted $R^2$ | 0.438 | 0.444 | 0.438 | 0.447 | 0.435 | 0.439 |
| N | 309,685 | 297,167 | 309,685 | 297,167 | 309,685 | 297,167 |

*Note:* The number in parenthesis is the t-statistic. *CRISIS* refers to the coefficient on the CRISIS dummy (the effect for the *omitted* medium cost group). Standard errors are corrected for clustering by time. */**/*** indicates statistical significance at the 10, 5 and 1 % level.

## Table 6: Excess comovements among size-sorted groups

| Factor | DAILY | WEEKLY | MONTHLY |
|---|---|---|---|
| Large Cap | 0.607 | 0.527 | 0.440 |
| | (536.08)*** | (153.97)*** | (13.75)*** |
| Mid Cap | 0.107 | 0.170 | 0.262 |
| | (21.11)*** | (19.47)*** | (19.02)*** |
| Small Cap | 0.172 | 0.167 | 0.111 |
| | (250.77)*** | (68.65)*** | (4.44)** |
| Adj. $R^2$ | 0.342 | 0.354 | 0.278 |
| N | 309,685 | 63,998 | 16,055 |

***Panel B: Average betas category-by-category (Weekly)***

| | Factor | | |
|---|---|---|---|
| Group | Small Cap | Mid Cap | Large Cap |
| Large Cap | 0.029 | 0.021 | 0.830 |
| | (2.16) | ( 0.55) | (187.63)*** |
| Mid Cap | 0.114 | 0.533 | 0.213 |
| | (5.33)** | (30.71)*** | (5.80)** |
| Small Cap | 0.837 | -0.016 | -0.018 |
| | (284.67)*** | (0.03) | (0.05) |

*Note:* The number in parenthesis is the F-statistics ($H_0$: *average* coefficient = 0). Standard errors are corrected for clustering by time. */**/*** indicates statistical significance at the 10, 5 and 1 % level.

**Table 7: Excess comovements among value-growth sorted groups**

| Factor | DAILY | WEEKLY | MONTHLY |
|---|---|---|---|
| Blend | 0.319 | 0.418 | 0.546 |
| | (81.55)*** | (26.41)*** | (29.47)*** |
| Value | 0.382 | 0.298 | 0.061 |
| | (112.49)*** | (11.05)*** | ( 0.30) |
| Growth | 0.134 | 0.118 | 0.242 |
| | (23.49)*** | ( 8.29)*** | ( 7.65)** |
| Adj. R2 | 0.331 | 0.340 | 0.306 |
| N | 309,685 | 63,998 | 16,055 |

***Panel B: Average betas category-by-category (Weekly)***

| | Factor | | |
|---|---|---|---|
| **Group** | **Blend** | **Value** | **Growth** |
| Blend | 0.267 | 0.118 | 0.388 |
| | (18.02)*** | ( 1.38) | (13.86)*** |
| Value | 0.411 | 0.507 | -0.070 |
| | (20.14)*** | (21.00)*** | ( 0.76) |
| Growth | 0.542 | 0.308 | 0.022 |
| | (13.58)*** | ( 4.25)** | ( 0.11) |

*Note:* The number in parenthesis is the F-statistics ($H_0$: *average* coefficient = 0). Standard errors are corrected for clustering by time. */**/*** indicates statistical significance at the 10, 5 and 1 % level.

**Table 8: Excess comovements among liquidity-sorted groups**

| Factor | DAILY | WEEKLY | MONTHLY |
|---|---|---|---|
| Low ILLIQ | 0.298 | 0.285 | 0.188 |
| | (94.70)*** | (17.10)*** | (3.87)** |
| Med ILLIQ | 0.306 | 0.278 | 0.337 |
| | (42.70)*** | (7.31)** | ( 7.41)*** |
| High ILLIQ | 0.252 | 0.269 | 0.296 |
| | (77.57)*** | (21.66)*** | (19.44)*** |
| Adj. R2 | 0.334 | 0.321 | 0.286 |
| N | 308,542 | 63,723 | 16,000 |
| Low QSPR | 0.401 | 0.266 | 0.267 |
| | (101.89)*** | (17.83)*** | (5.55)** |
| Med QSPR | 0.244 | 0.249 | 0.255 |
| | (36.39)*** | (12.21)*** | (4.60)** |
| High QSPR | 0.224 | 0.306 | 0.299 |
| | (101.00)*** | (42.79)*** | (24.89)*** |
| Adj. R2 | 0.321 | 0.331 | 0.276 |
| N | 308,542 | 63,723 | 16,000 |

***Panel B: Average betas category-by-category (Weekly)***

| | Interaction term | | | |
|---|---|---|---|---|
| **Factor** | **Low ILLIQ** | **High ILLIQ** | **Low QSPR** | **High QSPR** |
| Low ILLIQ | 0.2929 | -0.1477 | 0.3064 | -0.1686 |
| | (3.42)*** | (3.51)*** | (5.58)*** | (3.03)*** |
| Med ILLIQ | 0.0314 | -0.1237 | -0.1401 | -0.1501 |
| | (0.26) | (1.93)* | (2.86)*** | (2.83)*** |
| High ILLIQ | -0.0704 | 0.2651 | -0.1548 | 0.2199 |
| | (0.69) | (4.21)*** | (4.49)*** | (2.86)*** |
| Adjusted R2 | 0.322 | | 0.332 | |
| *N* | 63,723 | | 63,723 | |

*Note:* The number in parenthesis is the F-statistics ($H_0$: *average* coefficient = 0) in Panel A. T-statistics are reported in Panel B. Standard errors are corrected for clustering by time. */**/*** indicates statistical significance at the 10, 5 and 1 % level.

**Appendix 1: Variable definitions**

| Variable | Description | Definition |
|---|---|---|
| QSPR | Monthly average of closing quoted spread. | $QSPR_{i,m} = \frac{1}{D_{i,m}} \sum_{i=1}^{D_{i,m}} \left( P_{ASK,i} - P_{BID,i} \right) / \left( 0.5 P_{ASK,i} + 0.5 P_{BID,i} \right)$ |
| ILLIQ | Amihud's Illiquidity ratio capped at 30 % as in Acharya and Pedersen (2005). | $ILLIQ_{i,m} = \frac{1}{D_{i,m}} \sum_{i=1}^{D_{i,m}} \left| R_{i,m,d}^{RTF} \right| / VOLD_{i,m,d}$ where $D_{i,m}$ is the number of valid days in month m for stock *i*, *R* is the percentage return of ETF *i* in month *d*, day *d*. VOLD is the dollar volume for ETF *i* in month *m*, day *d*. |
| ID. VOL. | Idiosyncratic volatility, defined as in Baruch *et al.* (2010) | $ETF_{i,t}^{e} = a_i + \sum_{j=1}^{3} \beta_{i,j}^{US} R_{US,t} + \sum_{j=1}^{3} \beta_{i,j}^{UND} R_{UND,t} + e_{i,t}$ ID.VOL. is the monthly Std. Dev. of $e_{i,t}$. |