Journal of Banking and Finance 70 (2016) 118–136

Contents lists available at ScienceDirect

Journal of Banking and Finance

journal homepage: www.elsevier.com/locate/jbankfin

Systematic limited arbitrage and the cross-section of stock returns: Evidence from exchange traded funds

R. Jared DeLisle[a,1], Brian C. McTier[b,2], Adam R. Smedema[c,*]

[a] Utah State University, Department of Economics and Finance, Jon M. Huntsman School of Business, Logan, UT 84322, United States
[b] Washington State University, Department of Finance and Management Science, Carson College of Business, Vancouver, WA 98686, United States
[c] Tim Williams Fellow, University of Northern Iowa, Department of Finance, College of Business Administration, Cedar Falls, IA 50614, United States

ARTICLE INFO

Article history:
Received 7 March 2015
Accepted 5 June 2016
Available online 22 June 2016

JEL:
G12

Keywords:
Asset pricing
Factor model
Limits of arbitrage
Systematic
Risk
Arbitrage risk

ABSTRACT

We propose a parsimonious, comprehensive proxy for innovations in limited arbitrage: innovations in ETFs' premium. Consistent with a common component, we confirm limited arbitrage factors, LAFs, constructed from ETFs' premium innovations spanning four asset classes are correlated. Further, we find that equity LAFs are negatively priced in the cross-section of stock returns. Our pricing tests also confirm that LAFs provide pricing information beyond well-known limits of arbitrage: illiquidity and idiosyncratic volatility. Overall, our findings suggest that limited arbitrage risk is priced and LAF is a relevant risk-factor.

1. Introduction

As impediments to arbitrage and arbitrageurs' behavior critically underlie asset pricing, market efficiency depends on the ability of arbitrageurs to remove pricing errors. Previous research (Shleifer and Summers, 1990; Barberis and Thaler, 2003; Pontiff, 2006; Brav, Heaton, and Li, 2010) identify two theoretical limits to arbitrage: idiosyncratic volatility and illiquidity. Several studies (Ali, Hwang, and Trombley, 2003; Mashruwala, Rajgopal, and Shevlin, 2006; McLean, 2010; Stambaugh, Yu, and Yuan, 2015) analyze the cross-sectional implications of these limits.[3] Moreover, aggregate measures of idiosyncratic volatility (Brandt, Brav, Graham, and Kumar, 2010) and market illiquidity (Pastor and Stambaugh, 2003) have significant, systematic variation; they seem to not only vary cross-sectionally across assets, but also vary systematically over time.

Such systematic variation suggests limited arbitrage market states where overall mispricing in the market is high; for example, more assets will be mispriced when markets are, overall, less liquid than when overall market liquidity is high. Extant literature is supportive of the existence of limited arbitrage market states and provides some mechanisms for variation of such states. Deuskar (2007) models a market where increased volatility leads to less liquidity which in turn leads to additional volatility. Brunnermeier and Pedersen (2009) link market liquidity to the funding liquidity of arbitrageurs and this linkage could lead to increased market volatility and liquidity spirals. Mitchell and Pulvino (2012) show that after October 2008, several arbitrage funds simultaneously lost access to external financing. Hu, Pan, and Wang (2013) argue that the overall amount of capital available to arbitrageurs varies over time. Ang, Papanikolaou, and Westerfield (2014) predict that the average investor would pay a premium to hedge limited arbitrage resulting from the possibility of a liquidity crisis.

In this study, we begin with the premise that, while past literature finds that shocks to aggregate illiquidity and volatility are related to equity-market arbitrage, they are not likely the only factors limiting arbitrage during a given period nor are their impacts likely the same across periods or asset classes. This leads us to take a different approach in measuring innovations in and examining the impact of market-wide (systematic) limited arbitrage by exploiting the pricing of exchange-traded funds (ETFs). As ETF sponsors intend for ETFs to be close substitutes for the underlying assets, ETFs are designed so that arbitrageurs can capitalize on premium

* Corresponding author. Tel.: +1 319 273 6067; fax: +1 319 273 2922.
E-mail addresses: jared.delisle@usu.edu (R.J. DeLisle), b.mctier@wsu.edu (B.C. McTier), adam.smedema@uni.edu (A.R. Smedema).
[1] Tel.: +1 435 797 0885; fax: +1 435 797 2701.
[2] Tel.: +1 360 546 9516; fax: +1 360 546 9037.
[3] We use the term illiquidity to refer to frictions and costs in implementing a trade such as trading and borrowing costs.

http://dx.doi.org/10.1016/j.jbankfin.2016.06.006

deviations from zero at minimal cost in order to minimize tracking error. Therefore, we propose that the magnitude of changes in the premium on a given ETF would predominately reflect the net effect of changes of factors that systematically limit arbitrage of mispricing across multiple asset classes as well as mispricing specific to the ETF and underlying asset class. For example, considering the SPY ETF and the underlying S&P500 index, we posit that larger movements in SPY ETF premiums result from more limited arbitrage arising from many systematic factors that impact equity traders including, but not limited to, volatility and illiquidity.[4] Thus, instead of attempting to identify all possible factors that lead to limited arbitrage states, we consider the magnitude of changes in ETF premiums as a parsimonious measure that captures the relative inability of arbitrageurs to correct mispricing.

Further, consistent with past literature that finds that aggregate factors formed from idiosyncratic volatility, illiquidity, and other cross-sectional characteristics that limit arbitrage are priced in the cross-section of stocks, we expect that our measures formed from equity ETFs, capturing the net effects of limits of arbitrage that affect equity market traders, will be useful in pricing stocks. Consequently, the intent of this study is to both establish that ETF premium movements are *comprehensive* measures of innovations in systematic limited arbitrage and that limited arbitrage risk reflected in factors formed from equity ETFs is priced in the cross-section of stocks.

We calculate limited arbitrage proxies (LAPs) as the unsigned change in the ETF premium. We use the magnitude of the change because shocks to the ETF's underlying fundamental value can generate both overpricing and underpricing of the ETF and, therefore, the extent of the inability to arbitrage would be captured by the lack of (arbitrage-related) restraint on premium movements.[5] We construct tradeable limited arbitrage risk factors (LAFs) as zero-cost portfolios that are long the 20% of stocks with the most positive covariance with a particular LAP and short the 20% of stocks whose returns have the most negative covariance with that LAP. We construct seven LAFs from the mispricing of six ETFs that span four asset classes, S&P 500 (ticker: SPY), Dow 30 (DIA), Russell 3000 (IWV), U.S. Treasury bonds (IEF), Euros (FXE), gold (GLD), and the first principal component of the LAPs of the three equity ETFs. As, per our setup, higher LAP proxies for worsening arbitrage states; stocks with more positive covariance with LAP have relatively higher returns when arbitrage is worsening and stocks with more negative covariance with LAP have relatively lower returns. Since we construct our LAP as the return on the high portfolio minus the return on the low portfolio and, further, find that the average stock's return is negatively related to (has a lower covariance with) LAP, stocks in the low portfolio could be considered "riskier" or more arbitrage sensitive as their returns are lowest when LAP is high.

If our factors reflect systematic changes in the ability to perform arbitrage, then the arbitrage state variables (LAPs) and tradable risk factors (LAFs) constructed from ETFs from different assets classes should be correlated. To begin our study, we confirm this expectation by finding that the LAPs and LAFs constructed from the three equity ETFs are significantly correlated with each other at the 1% level. Similarly, the equity LAPs and LAFs are significantly correlated with the bond LAF. The currency and gold LAFs are also significantly correlated with the other LAFs. All of the correlations among the LAFs are positive. Together, these findings support the hypothesis that the changes in ETF premiums commonly capture changes in market-wide limited arbitrage as LAFs across several different asset classes are all positively and significantly correlated.

In the second part of our study, we test if limited arbitrage risk is priced using the three equity LAFs and their principle component. We limit our asset pricing tests to the relationship between equity ETF mispricing and equity prices for two reasons. First, as our test assets are equities, we expect the equity *LAFs* to more precisely capture innovations in limited arbitrage that impact equity market traders. Second, since the non-equity ETFs have shorter time-series, limiting our priced risk tests to equities reduces concerns that our findings are driven by the impacts of average negative news over the period of our study. Thus, though we find evidence for a common component of arbitrage risk in ETFs spanning multiple asset classes, the scope of our asset pricing tests limits us to forming conclusions about whether equity ETFs contain information about arbitrage risk that is priced in stocks.

We define limited arbitrage risk as a stock's sensitivity to (loadings on) a LAF and, consistent with extant literature regarding specific limits to arbitrage, we expect limited arbitrage risk to be negatively priced: stocks with more negative loadings on (covariance with) our limited arbitrage factor will require higher returns.[6] Our intuition follows from Pastor and Stambaugh's (2003) model where the representative investor is likely to liquidate stocks whose prices have fallen and trading costs systematically vary over time.[7] Stocks whose prices fall when trading costs rise will require costlier liquidation. Pastor and Stambaugh (2003) consider trading costs that vary with overall market liquidity. We argue, similarly, that potential trading costs vary with the level of systematic limited arbitrage (which includes, among other things, the overall liquidity of the market). Stocks that negatively load on our LAFs have prices that fall when arbitrage is systematically limited. Pastor and Stambaugh's (2003) investor will have to liquidate these stocks when mispricing, a potential additional trading cost, is most likely. Stocks with positive loadings on our limited arbitrage factor would be liquidated when mispricing is least likely and the potential trading costs from this mispricing are at a minimum. Hence, the in-

[4] Consider two shocks: a systematic funding shock that limits the ability of arbitrageurs to correct mispricing and new information that alters the fundamental value of a subset of S&P 500 index stocks but does not contain information about arbitrage. Assume that the new information, in the aggregate, has negative implications for stock prices and is immediately and efficiently reflected in the relevant individual stocks (note: the intuition of this example is unchanged if the assumption is that the information is reflected first in the price of the S&P 500 index ETF, SPY). If the drop in the value of the S&P 500 index was not immediately reflected in the price of SPY, then arbitrageurs would buy baskets of S&P 500 stocks and simultaneously contract to sell the basket to the Sponsor of the SPY ETF, capturing the difference. The cost of the transaction to the arbitrageur (SPY Prospectus, 2011) would include a $3000 (maximum) trustee fee, regardless of the number of baskets sold back, and, if the arbitrageur utilizes margin on the long-side or requires margin on the short side, slight noise trader risk from holding the stock position for 3 days before settlement with the sponsor. Overall, as long as the arbitrageur is able to hold the individual stock positions and withstand price movements over the 3 day settlement period, (s) he will capture the premium.

[5] To reduce concerns that our results are driven by mean reversion, we test and find that our SPY LAP (our longest time-series) is stationary and, further, do not find any statistically significant AR or MA components performing an ARMA(3, 3) regression with the SPY LAP. Additionally, we control for short-term reversals and momentum in our empirical tests.

[6] Our prediction is consistent with defining "risky" stocks as stocks with relatively more negative covariance with LAP (discussed above). While the pricing of limited arbitrage risk has not been explicitly examined in the literature, the pricing of other risks related to it (i.e., market volatility and illiquidity) strongly implies the negative price. Pastor and Stambaugh (2003) show that a factor that captures aggregate liquidity (illiquidity) is positively (negatively) priced in stocks. Similarly, building on Merton's (1973) intertemporal capital asset pricing model (ICAPM), Ang, Hodrick, Xing, and Zhang (2006) find that aggregate market volatility (measured with changes in VIX) is negatively priced and conclude that investors, hedging against a worsening investment opportunity set, will pay a premium to hold assets that are more sensitive to increasing volatility. Peterson and Smedema (2013) model aggregate idiosyncratic volatility (AIV) as cumulative across stocks, and find that AIV is negatively priced in the cross-section of stock returns.

[7] In Pastor and Stambaugh (2003), investors may need to liquidate stocks whose prices have fallen because of margin calls; however, there are other reasons for this liquidation including window-dressing (Lakonishok, Shleifer, Thaler, and Vishny, 1991).

vestor will accept a lower average return for stocks with more positive loadings on LAF and require a higher average return for stocks with more negative loadings.

Consistent with our expectation, we find a negative price of arbitrage risk. The average returns of rank portfolios formed from loadings (betas) on LAP decreases as the rank of the portfolio increases. Our LAFs generate negative average returns and factor-adjusted alphas. Zero-cost portfolios (LAFs) long the 20% of stocks with the highest (most positive) LAP betas and short the 20% of stocks with the lowest (most negative) LAP betas earn an economically and statistically significant average return of -0.28% per week when using the premiums of the S&P 500 ETF, SPY, to construct the LAF. Alphas from regressions of SPY LAF on Fama and French (1993) three-factor and Carhart (1997) four-factor models are a similar magnitude and statistical significance. These average returns and alphas are robust to both equal and value weighting the rank portfolios used to construct the LAF.

An alternative interpretation is that movements in premiums are not driven by the extent that arbitrage is worsening, but rather by demand shocks to stock prices and mean reversal from subsequent arbitrage trades. To examine this possibility, we additionally control for short-term reversals (Jegadeesh, 1990) and changes in non-fundamental demand with investor sentiment (Baker and Wurgler, 2006), innovations in the VIX (Ang, et al., 2006), and style momentum (Teo and Woo, 2004; Broman, 2015).

We find that the returns and alphas of the zero-cost portfolios are robust to inclusion of these additional controls. Further, the alphas are economically and statistically significant when controlling for aggregate idiosyncratic volatility and market illiquidity, consistent with the notion that the LAFs capture information about limited arbitrage beyond volatility and illiquidity. The portfolio results are similar when we construct the LAPs with the premium innovations of the Dow 30 ETF (DIA) and Russell 3000 ETF (IWV). When we use the principal component of the three equity ETF LAPs, we again find very strong evidence of pricing.

In addition to rank portfolios, we perform several additional tests. We estimate the price of arbitrage risk using Fama and MacBeth (1973) regressions. Regressing the returns of test portfolios and individual stocks on the stock's or portfolios' LAF betas (second stage) yields statistically significant negative coefficients (gammas) on LAF and is robust to controlling for additional factors including the aggregate idiosyncratic volatility and market illiquidity factors. We also perform rank portfolio tests where we first eliminate dispersion in sensitivity to aggregate idiosyncratic volatility and market illiquidity and the results remain robust. The Fama and MacBeth tests and the additional rank portfolio tests are consistent with LAFs capturing innovations in limited arbitrage driven by factors that limit arbitrage in the equities market, capturing information about aggregate idiosyncratic volatility and market illiquidity as well as information beyond them.[8] Additionally, we confirm our results are robust to exclusion of the period of the Great Recession and, thus, not driven solely by financial crises. We also compare the average characteristics of the high and low rank portfolios. We find that the low portfolio holds lower priced, smaller, and less liquid stocks with greater idiosyncratic risk consistent with our finding that these stocks have relatively lower returns when arbitrage is worsening (LAF is higher).

Overall, our findings suggest that, in general, ETF LAPs and LAFs from multiple asset classes reflect innovations in systematic limited arbitrage, that equity LAFs have a negative price of risk in the cross-section of stock returns, and that the equity LAFs are parsimonious risk factors that carry both information from factors that limit arbitrage of ETF and underlying equity mispricing as well as information beyond each individual limit to arbitrage.

Our paper proceeds as follows. Section 2 discusses relevant literature. Section 3 provides a framework for considering ETFs in the measurement of limits to arbitrage and describes our variable construction. Section 4 provides our testable hypotheses. Section 5 describes our data and controls; and Section 6 reports results from correlation tests among our LAF, LAP, and controls. Section 7 examines the cross-sectional pricing of our measures of limited arbitrage risk; Section 8 presents robustness checks and supplemental analysis; and Section 9 concludes.

2. Relevant literature

The finding of a priced systematic limited arbitrage factor contributes to three literatures. First, we contribute to the ETF literature by showing that premium movements across different ETF asset classes are correlated. Broman (2015) finds that correlated noise-trader demand for size and value investment styles generates a common factor in ETF premium changes. We include controls for style and confirm that our finding of a negative price of risk with our equity LAFs remains in our priced-risk tests. Further, Broman (2016) shows that ETF returns are affected by a common country-specific source of misvaluation that reflects the correlated non-fundamental demand of local investor clientele. We expand this by showing a common component extends to ETFs of different asset classes and that this common component has asset pricing implications.

Second, we contribute to the arbitrage literature with empirical evidence of the systematic nature of limited arbitrage. Third, we add to recent studies of financial contagion, dislocation, and arbitrage. Hu, Pan, and Wang (2013) use the mispricing of treasury securities as a measure of aggregate arbitrage activity and construct a factor to capture this “noise.” They show that hedge funds whose returns load negatively on this factor outperform hedge funds whose returns load positively on this factor. Pasquariello (2014) argues that periods of stress in financial markets should lead to bad market states. Using international measures of mispricing rather than domestic, he finds, similar to our findings, that stocks' ability to hedge these states explain the cross-section of returns. Our approach both complements and differs from Hu, Pan, and Wang (2013) and Pasquariello (2014); LAFs are not designed to capture the level of mispricing, but rather, by measuring the overall movements in premiums, we attempt to capture a measure of the overall “looseness” of the market or inability to perform arbitrage. When the ability to arbitrage is worsening, we expect premiums to be bounded by wider thresholds than when arbitrage is becoming more effective leading to larger innovations in premiums.

Acharya, Shin, and Yorulmazer (2013) argue that the trade-off between maintaining liquidity to take advantage of fire sale prices and the opportunity costs of reserve liquidity can lead to limited arbitrage across unrelated assets. We also find that limited arbitrage spans multiple assets and asset classes. However, our concern is not each individual mechanism that limits arbitrage, but rather the extent to which arbitrage is limited across the market. Thus, our measure would not only reflect the impacts of the idiosyncratic volatility, illiquidity, financial contagion, and fi-

[8] We propose that LAFs are a comprehensive proxy for innovations in systematic limited arbitrage. As such, this measure not only contains information about limits to arbitrage for the specific ETF and underlying asset, but also contains information about limits to arbitrage that affect multiple assets. Past literature identifies many candidates that limit arbitrage including Barberis and Thaler (2003) limits and other factors based on volatility (see Adrian and Rosenberg, 2008; Duarte, Kamara, Siegel, and Sun, 2014; Herskovic, Kelly, Lustig, and van Nieuwerburgh, 2014). Further, there are certainly many as of yet unidentified factors. For example, Chou, Huang, and Yang (2013) suggest that the stock turnover anomaly is related to arbitrage risk. As it would be impractical to exhaustively control for each identified factor and impossible to control for unknown factors, to perform our tests, we narrow the controls we use to control for known limits of arbitrage to two commonly accepted limits to arbitrage: idiosyncratic volatility and illiquidity (e.g. Pontiff, 2006).

nancial dislocations during the financial crisis, but would reflect systematic limited arbitrage that impacts multiple asset classes as well as limits to arbitrage specifically impacting the ETF or its underlying asset class.

3. Measuring limited arbitrage

3.1. Exchange traded funds

To identify innovations in the limited arbitrage market state, we utilize the unsigned innovation in premiums of three equity ETFs (SPY, DIA, and IWV) and three non-equity ETFs (IEF, FXE, and GLD). ETFs are constructed to allow investors to trade large baskets of assets or to gain exposure to otherwise difficult to buy or short assets. The market price of the shares of the ETF is designed to track the value of the underlying as the fund sponsors allow for relatively easy arbitrage by large, sophisticated institutional investors ("authorized participants") who can create overvalued shares or redeem undervalued shares of an ETF when its price diverges from the underlying.[9] This "in-kind" creation and redemption of shares is deliberately designed to force the ETF premium or discount close to zero; thus, we view the ETF premium as an asset with a known, correct price of zero, but real-world impediments remain (Shleifer and Summers, 1990) and divergence from zero premiums may emerge.

3.2. Proxies for state variables

As our empirical measure of innovations in the limited arbitrage state, we calculate the absolute value of the change in the ETF premium. Thus, our expectation is that as arbitrage becomes more limited, the bounds between the prices of the ETF and the NAV will be greater resulting in larger premium swings. We calculate the premium $PREM_{it}$ for ETF i on week t as

$$PREM_{it} = \frac{ETF_{it} - NAV_{it}}{NAV_{it}} \tag{1}$$

where ETF_{it} and NAV_{it} are the price and NAV of ETF i, at the end of week t, respectively. Our proxy for the innovation in the limited arbitrage state, LAP, for ETF i week t, is

$$LAP_{it} = |\Delta PREM_{it}| \tag{2}$$

The changes in the premium are calculated as the Friday-over-Friday change. We use weekly data due to the tradeoff between noise from shorter periodicities (i.e., daily) and the sample size constraint of longer periodicities (i.e., monthly) because of the relatively short lengths of the time-series for ETFs. Also, compared to using daily returns, using weekly frequencies reduces concerns about nonsynchronous trading between ETFs and their underlying.

Our limited arbitrage proxies (LAPs) are superior to alternative statistical measures of limited arbitrage for two key reasons. First, since the current market price and the underlying value of the ETFs (net asset value, or NAV) are both observable, innovations in the premium are also observable. Therefore, we do not need to estimate an otherwise unobservable limit or construct an index of mispricing that must weight several different mispricing measures.[10] Instead, we can directly observe the magnitude of premium changes of these relatively easy to arbitrage assets and, by our hypothesis, market states in which impediments to arbitrage are increasing or decreasing. Second, we propose that innovations in the ETF premium are comprehensive measures of changes in systematically limited arbitrage. There are an unknown number of limits to arbitrage. Therefore, rather than theoretically identifying and estimating each limit individually, we posit that we are simply observing the total effect of the limits to arbitrage impacting traders of the ETF and underlying simultaneously. The parsimonious nature of our measure, then, should be superior to aggregate idiosyncratic volatility and market illiquidity in explaining the cross-section of stock returns since our measure includes not only these limits to arbitrage, but other equity market limits not yet documented or identified.

3.3. Limited arbitrage factors

We construct seven risk factors from the ETF LAPs using factor-mimicking portfolios to transform non-traded variables into tradable assets. At the beginning of each week, we estimate the following regression for each stock in our sample with the previous 13 weeks of data

$$r_{it} - rf_t = \beta_{i,MRP} \cdot MRP_t + \beta_{if} \cdot f_t + u_{it}, \quad t \in [-1, \ -13] \tag{3}$$

where $r_{it} - rf_t$ is stock i's return in excess of the risk free rate at week t, MRP_t is the week t market risk premium, f_t is the week t factor innovation, and u_{it} is random error. This procedure is similar to that used by Lewellen and Nagel (2006) who estimate quarterly betas using weekly data over the quarters. We include MRP_t controlling for other omitted relevant systematic effects. We then form an equal-weighted, zero-cost portfolio with week t=0 returns that is long the 20% of stocks with the highest β_{if} and short the 20% of stocks with the lowest β_{if}.[11] We repeat this process for SPY, DIA, IWV, IEF, FXE, GLD, and the principle component LAP.

3.4. Additional state variables and factors

Similarly to our sample of ETFs, we construct limited arbitrage proxies for aggregate idiosyncratic volatility and market illiquidity.[12] We construct innovations in aggregate idiosyncratic volatility similarly to Goyal and Santa-Clara (2003). Using Center for Research in Security Prices (CRSP) daily data, we calculate aggregate idiosyncratic volatility, AIV_t, for week t each Friday:

$$AIV_t = \frac{1}{n}\sum_{i=1}^{n}(r_{id} - \bar{r}_d)^2 \tag{4}$$

where r_{id} is the Friday return for stock i and $\bar{r}_d$ is the equal-weighted average return for the day. AIV LAP is the innovation in the weekly AIV.

We construct innovations in market-wide illiquidity from value-weighted aggregate illiquidity (Amihud, 2002) similar to Watanabe and Watanabe (2008). The Friday, week t illiquidity across i stocks is

$$MILLIQ_t = 10^4 \cdot \sum\nolimits_{i=1}^{n} x_i \cdot \frac{|r_{it}|}{VOL_{it}} \tag{5}$$

where $|r_{it}|$ is the absolute value of the Friday return for stock i, VOL_{it} is the daily dollar volume for stock i, and x_i is the market-capitalization weight of firm i at the close Thursday.

Similar to the construction of our LAFs (Section 3.3), we fit Eq. (3) with the AIV and $MILLIQ$ state variable innovations (f_t) and construct corresponding risk factors using value-weighted factor-mimicking portfolios.

[9] See Broman (2015) for discussion of the mechanics of "in-kind" ETF share creation and redemption.

[10] Ali et al. (2003) and Stambaugh et al. (2015) estimate limited arbitrage with an estimate of idiosyncratic volatility. Pasquariello (2014) constructs a measure of market dislocation from violations of well-known arbitrage parity conditions.

[11] We examine using value-weighted portfolios and different formation periods to form our LAFs in Section 7.2.

[12] We construct our liquidity factor rather than use Pastor and Stamabugh (2003) liquidity factor because we have insufficient degrees of freedom to estimate their liquidity prediction model as weekly. Further, there is not an obvious way to decompose the monthly index into weekly innovations.

4. Empirical hypotheses

To empirically investigate the existence and cross-sectional asset pricing of limited arbitrage states and that LAFs capture innovations in these states, we derive four testable hypotheses. First, with time-varying states of limited arbitrage, individual asset prices will collectively deviate from their fundamental values within decreasing and increasing thresholds and, thus, the magnitudes of premium changes of ETFs with different underlying assets will be correlated over time. We focus on innovations in the premium because, in a limited arbitrage state, larger premium swings may occur, but not necessarily in the same direction for all assets. As we interpret correlated premium innovations as evidence of common proxies for limited arbitrage, we expect the returns of zero-cost portfolios that are constructed to convert these non-traded variables into tradable limited arbitrage risk factors (LAF) are correlated. Our first hypothesis (stated in alternate form) follows:

H_1. *Tradable ETF LAFs are positively correlated.*

Our second hypothesis follows from our prediction that arbitrage risk will be negatively priced. Consistent with the economic intuition of Ang et al. (2006), investors will pay a premium and accept lower average returns for stocks that have relatively high payoffs as arbitrage becomes systematically limited (our high rank portfolio) because these stocks hedge investors' portfolio exposure to limited arbitrage states. Similarly, consistent with Pastor and Stambaugh (2003), investors will require a higher return to hold stocks that payoff relatively poorly when arbitrage is becoming more limited (our low rank portfolio) as their marginal utility is higher in these states and forced liquidation of the poorly performing stocks is costlier. Therefore, we propose that stocks with more positive loadings on LAP and LAF should have lower expected returns than stocks with more negative loadings ceteris paribus. Our second hypothesis (stated in alternative form) follows:

H_2. *Stocks whose returns load more positively on LAP and LAFs have lower expected returns than stocks whose returns load more negatively.*

Our third and fourth hypotheses follow from the notion that time-varying limited arbitrage is driven by the cumulative impact of specific time-varying limits to arbitrage. Existing empirical studies of the asset pricing implications of limits to arbitrage focus on cross-sectional differences in the pricing of stocks. In general, these studies show that asset pricing model anomalies such as the value effect (Ali et al., 2003), sentiment effect (Baker and Wurgler, 2006), accruals anomaly (Mashruwala et al., 2006), and momentum effect (McLean, 2010) are stronger in stocks with characteristics that limit arbitrage. However, the well-known limits to arbitrage, aggregate idiosyncratic volatility and market illiquidity (Pontiff, 2006), considered in this study also exhibit significant time-series variation; market states exist in which stocks have more firm-specific risk and markets are more illiquid.[13] Intuitively, as arbitrage becomes more limited, all assets are relatively more prone to mispricing. As such, we propose that our measure is a parsimonious proxy for overall movements in limited arbitrage and that these overall movements are driven by both known and unknown limits to arbitrage. Our third and fourth hypotheses follow:

H_3. *LAFs contain information about individual (specific) limits to arbitrage.*

H_4. *LAFs contain information about systematic limited arbitrage beyond individual limits to arbitrage.*

5. Data

5.1. Sample of ETFs

Consistent with our hypothesis that states of limited arbitrage affect multiple assets and asset classes, we choose our sample of ETFs to span several broad-based asset classes: major equity market indices, bonds, currency, and gold. We choose ETFs that track broad-based market indices and asset classes to mitigate concerns that our findings reflect the effects of stale pricing or style-based trading (Broman, 2015). Furthermore, as ETFs were not prevalent prior to the year 2000, to reduce noise or the impact of other factors, we choose a sample of ETFs that are actively traded and have a reasonably long time-series. Our equity ETFs are the SPDR S&P500 (trading symbol: SPY), the SPDR Dow Jones Industrial Average (DIA), and the iShares Russell 3000 (IWV). Our bond ETF is the iShares 7–10 Year Treasury bond (IEF), a portfolio of US Treasury Notes constructed to closely match the Barclays US 7–10 Year Treasury Index. Our currency ETF is the CurrencyShares Euro Trust (FXE), a portfolio of Euros, and our gold ETF is the SPDR gold trust (GLD), a portfolio of gold.

By construction, ETF premiums and discounts are sensitive to nonsynchronous trading. If ETF and NAV prices are measured at different points in time, there is a possibility of miscalculating the premium. The NAV of the equity ETFs and IEF are measured at the same time as they are time-stamped at 4:00PM EST on Bloomberg which corresponds with the time-stamp on ETF prices from CRSP. However, the NAV for FXE and GLD do not coincide with the ETF prices reported in CRSP; the NAV for GLD is set as of the London AM fix and the NAV for FXE is set at the close of U.S. currency futures markets. Therefore, to reduce the impact of nonsynchronous trading, we collect ETF prices for GLD and FXE from the NYSE Trade and Quote system (TAQ) by choosing the intraday prices that are closest to the time-stamp of the reported NAV.[14]

Since the ETFs in our sample came into existence at different times, our series for each is a different length. The equity ETFs SPY, DIA, and IWV began in 1993, 1998, and 2000, respectively. IEF began in 2002, and FXE and GLD began in 2005 and 2004, respectively. Our data for SPY, DIA, IWV, IEF all end on 31 December 2011, which provides us 986, 727, 604, and 492 weekly observations, respectively. Our data for GLD and FXE end on 31 December 2010, which provides 320 and 264 weekly observations, respectively.[15] Summary statistics for the weekly ETF returns are reported in Table 1, Panel (A).[16]

[13] The time-series variation in aggregate idiosyncratic volatility has been shown in Campbell, Lettau, Malkiel, and Xu (2001), Goyal and Santa-Clara (2003), Brandt, Brav, Graham, and Kumar (2010), Peterson and Smedema (2013), and Herskovic, et al. (2014). The time-series variation in market liquidity has been shown in Chordia, Roll, and Subrahmanyam (2000, 2011), Pastor and Stambaugh (2003), Gibson and Mougeot (2004), and Cao, Chen, Liang, and Lo (2013).

[14] Alternatively, instead of using ETF prices and NAVs as reported, they can be estimated following the peer group methodology of Broman (2015). This methodology is designed to minimize the impact of staleness or other non-economic aberrations in the quoted prices and NAVs. However, considering the tradeoff between using long enough time-series so that are results are not driven by average news over the period and less noise in the ETF premium innovations, we are less concerned about noise as it works against us finding that arbitrage risk is priced. As such, any statistical significance we find would be in spite of, not because of, this noise. We also use only actively traded ETFs to reduce concerns about noise (discussed in Sec. 4.1).

[15] Our TAQ data sample required to set the closing price of our GLD and FXE ETFs ends on 31 December 2010.

[16] For brevity, we omit discussion of the ETF returns for our sample, but we still report the results for completeness.

Table 1
Summary information.

	Obs.	Mean (%)	SD (%)	Min (%)	Max (%)	Beg. week
Panel A – ETF returns						
SPY	986	0.172	2.481	−19.790	13.290	1993-05
DIA	727	0.134	2.615	−18.850	12.530	1998-03
IWV	604	0.067	2.742	−17.170	12.170	2000-22
IEF	492	0.128	0.937	−2.822	3.395	2002-31
GLD	319	0.406	3.051	−13.580	15.340	2004-47
FXE	264	0.087	1.490	−5.466	5.598	2005-50
Panel B – proxies for state variables						
SPY_LAP	985	0.220	0.322	0.000	4.025	
DIA_LAP	726	0.205	0.342	0.000	4.523	
IWV_LAP	603	0.168	0.244	0.000	2.105	
IEF_LAP	491	0.095	0.110	0.000	1.259	
GLD_LAP	318	0.270	0.484	0.000	4.869	
FXE_LAP	263	0.108	0.121	0.000	0.794	
*PC*_LAP	603	0.319	0.393	0.017	3.522	
*AIV*_SV	986	−0.000	0.082	−0.864	0.914	
*MILLIQ*_SV	986	−0.000	0.024	−0.143	0.184	
*VIX*_SV	986	0.000	0.030	−0.192	0.248	
*SENT*_SV	934	−0.003	0.237	−0.823	1.054	
Panel C – tradable risk factor returns						
SPY_LAF	972	−0.272	1.567	−13.420	7.949	
DIA_LAF	713	−0.279	1.746	−12.720	8.138	
IWV_LAF	590	−0.217	1.368	−7.490	3.966	
IEF_LAF	478	−0.204	1.223	−4.828	5.316	
GLD_LAF	305	−0.215	1.384	−6.651	5.446	
FXE_LAF	250	−0.165	1.332	−5.143	4.729	
*PC*_LAF	590	−0.317	1.407	−7.210	5.142	
*AIV*_RF	973	−0.114	1.999	−11.070	10.850	
*MILLIQ*_RF	973	−0.019	1.994	−12.900	16.340	
MOM	973	0.133	2.422	−15.750	12.500	
STREV	973	0.446	2.253	−18.410	20.510	
*VIX*_RF	973	−0.033	2.012	−11.440	14.590	
*SENT*_RF	921	0.135	2.511	−22.510	10.710	

This table reports summary information for variables of interest in our study. Panel A reports summary information for weekly returns of ETFs: SPY, DIA, IWV, IEF, GLD, and FXE (described in Table 1). Panel B reports summary information for innovations in weekly limited arbitrage proxies (LAP) and state variables (SV), Friday to Friday flows. ETF LAP variables are formed as the unsigned innovation in the ETF's premium. PC is the first principal component of the SPY, DIA and IWV LAPs. *AIV*_SV is the weekly innovation in aggregate idiosyncratic volatility constructed similarly to Goyal and Santa-Clara (2003), *MILLIQ*_SV is the innovation in aggregate market illiquidity, constructed from the value-weighted average Amihud (2002) illiquidity measure similar to Watanabe and Watanabe (2008), and *VIX*_SV is the change of the VIX as reported on the CBOE. *SENT* is the monthly change in sentiment from Baker and Wurgler (2006), interpolated weekly. Panel C reports summary information for the weekly returns of tradable risk factors and controls. For our sample of ETFs and principal component, limited arbitrage risk factors (LAF) are constructed from LAPs with zero-cost factor mimicking portfolios. *AIV, MILLIQ, VIX*, and *SENT* risk factors (RF) are zero-cost portfolios similarly constructed from their weekly innovations. *MOM* and *STREV* are momentum and short-term reversal factors downloaded from Ken French's website. FXE, GLD, and *SENT* data are collected through December 31, 2010 and the remainder is collected through December 31, 2011.

5.2. Compustat and CRSP data

To perform our pricing tests and construct state variables, limited arbitrage risk factors, and test portfolios, we form a sample of ordinary common stocks (CRSP share codes 10 and 11) that trade on the NYSE, AMEX, or NASDAQ (CRSP Exchange codes 1, 2, and 3) and have a price of at least $5. Book equity is computed from Compustat data. Firms with missing or negative book equity are omitted from the sample.

5.3. Proxies for state variables

As described in Sections 3.2 and 3.4, we construct limited arbitrage proxies (LAPs) from our six ETFs and state variable flows for aggregate idiosyncratic volatility (*AIV*) and market illiquidity (*MILLIQ*). We report summary statistics for the ETF LAPs and the principle component of the three equity LAPs in Rows (1)–(7) in Panel (B) of Table 1. Mean LAP for all six ETFs range from 0.10% (IEF) to 0.27% (GLD). The standard deviations range from 0.11% to 0.48% suggesting that our proxies would capture significant variation in innovations to systematic limited arbitrage. Summary statistics for *AIV* LAP *and MILLIQ* LAP are reported in Table 1, Panel (B), Rows (8) and (9).

5.4. Systematic demand shocks

A concern is that the magnitude of premium innovations (LAPs) not only captures a worsening state of limited arbitrage, but also captures systematic shocks that selectively impact either the ETF or the underlying assets.[17] To control for this possibility, we consider two additional factors: market volatility and investor sentiment. We construct state variable proxies to control for "investor fear" (see Whaley, 2000 for a discussion of the VIX index as the investor fear gauge) or, alternatively, changing investor uncertainty about market conditions, with innovations in the VIX index

[17] We thank an anonymous reviewer for bringing this point to us.

Table 2
ETF information.

	SPY	DIA	IWV	IEF	FXE	GLD
Cumulative DVol (million)	$331.082	$28.918	$187	$113	$225	$6183
Number of assets	500	30	3000	15	1	1
Days to settle	3	3	3	3	3	3
Trustee fee	Min. $3000 or 0.1% of transaction	Min $1000 or 0.1% of transaction	$3000	$0	$500	Min. $2000 or 0.1% of transaction
Amihud illiquidity measure	0.0003	0.0049	0.3229	0.1656	0.1100	0.0106
Bid ask spread (%)	0.0567	0.1235	0.1887	0.7938	0.3062	0.4396

This table reports information for ETFs in the study: SPDR S&P 500 (SPY), SPDR Dow Jones Industrial Average (DIA), iShares Russell 3000 (IWV), iShares 7–10 Year Treasury Bond Fund (IEF), CurrencyShares Euro Trust (FXE), and SPDR Gold Trust (GLD). Trade information is collected from the NYSE Trade and Quote (TAQ) system and daily return and dollar volume is collected from the CRSP daily file over the period January 1, 2008 through March 31, 2008. The remainder of the information is collected from the fund's prospectus. Cumulative DVol is the cumulative dollar volume (in millions) of all of the trades during the sample period. Number of Assets is the number of assets to create a share (or basket) of the ETF; Days to Settle is the number of days between the date the purchase order from the authorized participant is accepted and settlement; and Trustee Fee is the fee charged to the authorized participant regardless of the number of creation units (baskets) contracted. Amihud Illiquidity Measure is the average over the period of the daily Amihud measure calculated as the absolute value of the ETF's daily return divided by the daily dollar volume (Amihud, 2002). Bid Ask Spread is the average over the period of the daily closing percentage bid-ask spread.

(Ang et al., 2006). We construct state variable proxies to control for changing investor sentiment with innovations in the Baker and Wurgler (2006) investor sentiment index. As described in Section 3.4, we construct corresponding risk factors, *VIX* and *SENT*, using value-weighted factor-mimicking portfolios. We obtain the VIX index levels from Bloomberg and the levels and innovations in the sentiment index (SENT) from Jeffrey Wurgler's website.[18] We linearly interpolate weekly innovations in *SENT* as the downloaded data is monthly. We report summary statistics for the *VIX* and *SENT* state variable innovations in Rows (12) and (13) in Panel C of Table 1.

5.5. Risk factors

As described in Sections 3.3, 3.4, 5.3 and 5.4 above, to perform our tests, we construct eleven risk factors using factor-mimicking portfolios to transform non-traded variables into tradable assets: seven ETF LAFs, aggregate idiosyncratic volatility (*AIV*), market illiquidity (*MILLIQ*), VIX, and sentiment (*SENT*). We also obtain Fama and French (1993) three-factors, Carhart (1997) momentum, and Jegadeesh (1990) short-term reversal from Ken French's website.[19] Inclusion of these controls is fairly standard in the asset pricing literature; however, we also include them to help rule out alternate explanations of what our LAFs capture. We include Fama and French (1993) three factors to control for time-variation in investor preferences for particular styles and strategies. As premium innovations are mean-reverting, we include the short-term reversal factor to control for our LAFs spuriously capturing short-term serial correlation in stock returns.

We report summary statistics of the LAFs (zero-cost portfolio returns) in the first seven rows of Panel C of Table 1. Consistent with Hypothesis H_2 and our expectation that limited arbitrage risk is priced, mean returns on all of our ETF LAFs are negative with an economically significant magnitude. Mean weekly return for SPY is −0.27% (−14% annualized), DIA is −0.28% (−14.5% annualized), IWV is −0.22% (−11% annualized), and IEF is −0.20% (−11% annualized). The mean weekly returns for GLD and FXE LAFs are −0.22% and −0.17%. The LAF from the principal common *PC* is the strongest with an average weekly return of −0.32%. The LAFs have standard deviations between 1.2% and 1.7% per week. We report summary statistics for the control risk factors (*AIV, MILLIQ, MOM, STREV, VIX*, and *SENT*) in the last six rows of Panel C of Table 1. We find, consistent with previous literature, the average returns on *AIV* (Peterson and Smedema, 2013), *MILLIQ* (Pastor and Stambaugh, 2003; Watanabe and Watanabe, 2008), and *VIX* (Ang et al., 2006) are negative and *SENT* has positive average returns (Baker and Wurlger, 2006).[20]

5.6. Additional ETF information

To provide preliminary evidence that premium innovations are predominantly driven by changes in systematic limited arbitrage, we examine the trading activity and cost of capturing arbitrage profits for each of our six ETFs; greater trading activity and lower costs would facilitate trades by large, institutional arbitrageurs.

Row (1) of Table 2 reports the cumulative dollar volume of trades for our sample of ETFs collected from the NYSE Trade and Quote (TAQ) system over the period 1 January 2008 through 31 March 2008.[21] Rows (2)–(5) of Table 2 consider the direct costs to an authorized participant of capturing an arbitrage opportunity through transactions with the fund's sponsor. Row (2), Number of Assets, reflects the complexity of the holdings required to create shares. Simultaneously acquiring multiple positions may generate significant fees for the authorized participant who may be unable to enter into positions at the desired prices and as fund sponsors essentially desire zero tracking errors, using a synthetic position to capitalize on ETF mispricing may be more difficult than simply matching the assets held by the fund. We measure complexity with the number of assets required to create shares. SPY and IWV have significantly higher complexity than the remaining ETFs as they require creating positions in the majority of 500 and 3000 stocks, respectively, while DIA requires creating positions in 30 stocks, IEF requires positions in multiple Treasury note futures, and FXE and GLD require only a single asset.[22]

Row (3) reports the number of days between the date the purchase order from the authorized participant is accepted and settlement. Greater days to settlement exposes the arbitrageur

[18] http://people.stern.nyu.edu/jwurgler/

[19] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[20] In untabulated results, we perform t-tests on the constructed risk factor controls (H-L zero-cost returns) and calculate the alphas from a CAPM and *FF3* model. We find that the returns and alphas of these factors have the correct sign; however, they lack statistical significance with the exception of the VIX factor from Ang, et al. (2006). Pastor and Stambaugh (2003), Baker and Wurgler (2006), and Peterson and Smedema (2013) all look at long-term horizons (three to five years), to estimate their factor loadings. However, our measures of arbitrage risk (LAFs) are weekly (due to constraints discussed in Section 3.2) using a 13 week formation period. Ang et al.'s (2006) VIX factor is, similar to our VIX factor, a short-term measure and they find a similar negative price of market volatility risk.

[21] The three-month period for our trade size subsample was selected to be representative of the entire time period of our study, i.e. near the mid-point of the sample and sufficiently later than the inception of FXE, the ETF with the shortest time-series.

[22] We say "majority" because the authorized participant can replace some, but not all securities in the creation unit with the security's cash equivalent.

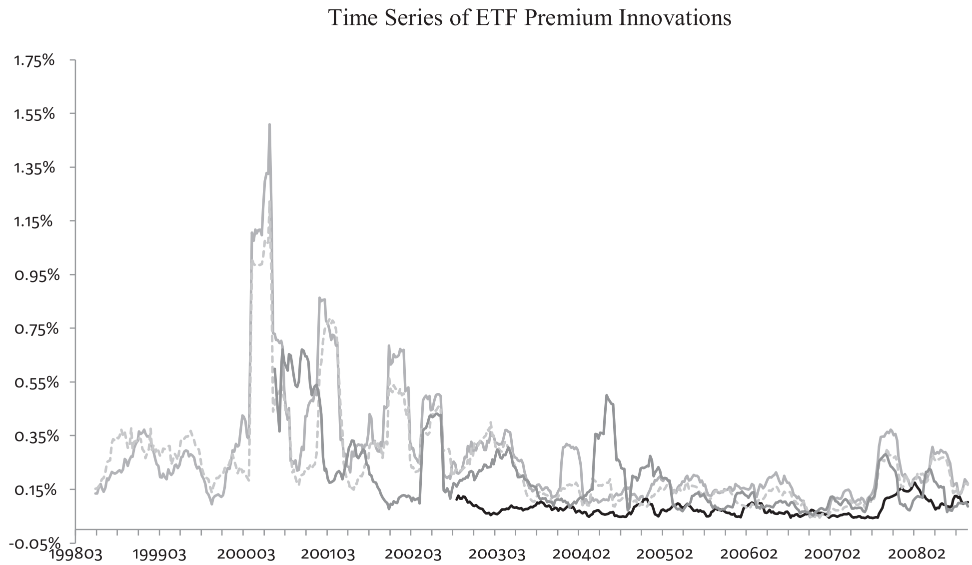

Fig. 1. plots the 10-week moving averages of the SPY (light grey line), DIA (dotted grey), IWV (dark grey), and IEF (black) LAPs.

utilizing margin to slight noise trader risk; however, profits are guaranteed as long as the arbitrageur can withstand intermediate market movements in the underlying asset. Row (4) reports the fee charged by the fund sponsor to enter into a purchase order. For all ETFs, the fee is a fixed maximum amount regardless of the number of creations units (baskets) contracted for in the purchase order and thus can be minimized with large purchase orders. Rows (5)–(6) report the average daily Amihud illiquidity measure (Amihud, 2002) and percentage bid-ask spread over the period. SPY and DIA have the lowest bid-ask spreads and impact of trading. IWV has the highest Amihud measure consistent with the number of assets required to create a basket. Overall, the trading costs of SPY and DIA appear minimal with the possible exception of SPY where the arbitrageur would be required to transact in 500 stocks to capture the arbitrage profits without incurring the additional risk of a synthetic position. Further, SPY, DIA, and GLD have relatively greater volume to support large arbitrageur trades. However, all of the non-equity ETFs have higher bid-ask spreads.

6. Correlation of LAPs and LAFs

In this section, we test Hypothesis H_1 by estimating the correlations among our seven ETF LAPs and LAFs. Fig. 1 plots the 10-week moving averages of the SPY (light grey line), DIA (dotted grey), IWV (dark grey), and IEF (black) LAPs. Overall, the LAPs appear to be correlated. The SPY and DIA LAPs largely move in lock-step throughout the time-series. The IWV LAP is much less volatile than the SPY and DIA LAPs, but still moves in a similar pattern. All three were above average during the fallout from the Tech bubble (2000 through 2001), at low levels over the period 2003 through 2006, and noticeably rose in 2007 with a large spike up during the Financial Crisis in late 2008. IEF, which is the shortest time series, was not traded during the fall out of the Tech bubble but its LAP does exhibit a spike during the Financial Crisis.

Table 3 reports pairwise correlations. In Panel A, we report pairwise correlations among ETF LAPs and proxies for state variables: *AIV, MILLIQ, VIX*, and *SENT*. In Panel B, we report correlations between the LAFs and tradeable risk factor controls: *AIV, MILLIQ, MOM, STREV, VIX*, and *SENT*.

In Panel A, consistent with a common component that reflects changes in systematic limited arbitrage (our H_1), the correlations among the equity ETF, equity principal component, and the bond ETF LAPs are positive and significant at the 1% level. The correlation between the LAPs of GLD and FXE is positive and significant at the 5% level; however the GLD and FXE LAPs are uncorrelated with the equity ETFs. Consistent with ETF premium innovations corresponding to states of market illiquidity and volatility, the SPY LAP is positively correlated with the *MILLIQ* and *VIX* state variables, and the DIA LAP is positively correlated with *AIV* state variable. SPY LAP is negatively correlated with the *SENT* state variables. The equity principal component LAP, which captures the commonality in the LAP of the equity ETFs, is significantly correlated with the *AIV* and *VIX* state variables in the expected directions. Overall, the findings in Panel A are consistent with commonality in the premium innovations of the equity ETF LAPs and this commonality is correlated with the *AIV, MILLIQ, VIX*, and *SENT* state variables in the expected directions.

In Panel B, again consistent with a common component that reflects changes in systematic limited arbitrage (our H_1), the correlations among all of the ETF LAFs are positive and significant at the 1% level. The pairwise correlations of the three equity LAFs range from 0.69 to 0.92, with the highest correlation between SPY and DIA. As expected, *AIV* and *VIX* are significantly positively correlated with each other, *MILLIQ*, and with the SPY and DIA LAF. *MILLIQ* is uncorrelated with the SPY and DIA LAFs.

7. Pricing of systematic arbitrage risk

7.1. Overview of cross-sectional tests

We examine if arbitrage risk is negatively priced in the cross-section of stock returns, our Hypothesis H_2. We then test to confirm that equity LAFs reflect information about well-known limits of arbitrage (Hypothesis H_3) as well as additional information about systematic limited arbitrage beyond these specific limits (Hypothesis H_4). As described above in Sections 1 and 4, we predict that stocks that perform relatively well when arbitrage becomes more systematically limited will be priced at a premium and earn lower average returns over time compared to stocks that perform relatively poorly when the ability to perform arbitrage is worsening. Coupled with our thesis that ETF LAPs and LAFs capture changes in systematic limited arbitrage, we predict that stocks that covary more positively with our LAPs and LAFs will have lower

Table 3
Pairwise correlations of state variable innovations and tradeable risk factors.

Panel A – proxies for state variables

	SPY_LAP	DIA_LAP	IWV_LAP	IEF_LAP	GLD_LAP	FXE_LAP	*PC_LAP*	*AIV_SV*	*MILLIQ_SV*	*VIX_SV*
DIA_LAP	0.83***									
IWV_LAP	0.34***	0.32***								
IEF_LAP	0.29***	0.15***	0.26***							
GLD_LAP	−0.03	−0.02	0.04	−0.05						
FXE_LAP	0.07	0.07	0.02	0.08	0.13**					
PC_LAP	0.93***	0.91***	0.57***	0.29***	−0.02	0.07				
AIV_SV	0.02	0.06*	−0.01	0.01	−0.00	0.06	0.11***			
MILLIQ_SV	0.06**	0.02	−0.01	0.02	0.01	0.02	0.06	0.12***		
VIX_SV	0.07**	0.05	0.03	0.16***	−0.09	−0.01	0.07*	0.12***	0.07**	
SENT_SV	−0.07**	0.04	−0.01	−0.09*	−0.01	−0.02	−0.17***	0.01	0.02	−0.03

Panel B – tradeable risk factors

	SPY_LAF	DIA_LAF	IWV_LAF	IEF_LAF	GLD_LAF	FXE_LAF	*PC_LAF*	*AIV_RF*	*MILLIQ_RF*	*MOM*	*STREV*	*VIX_RF*
DIA_LAF	0.92***											
IWV_LAF	0.69***	0.70***										
IEF_LAF	0.73***	0.72***	0.81***									
GLD_LAF	0.68***	0.67***	0.76***	0.80***								
FXE_LAF	0.71***	0.65***	0.75***	0.77***	0.79***							
PC_LAF	0.95***	0.95***	0.81***	0.78***	0.73***	0.75***						
AIV_RF	0.13***	0.11***	−0.06	−0.14***	−0.12**	−0.03	0.12***					
MILLIQ_RF	0.01	0.00	−0.16***	−0.05	−0.07	−0.01	−0.09**	0.11***				
MOM	0.42***	0.47***	0.47***	0.45***	0.55***	0.40***	0.40***	−0.13***	−0.20***			
STREV	−0.30***	−0.31***	−0.44***	−0.45***	−0.49***	−0.34***	−0.36***	0.05	0.19***	−0.23***		
VIX_RF	0.17***	0.16***	−0.09**	0.01	0.08	0.04	−0.04	0.09***	0.07**	0.12***	0.11***	
SENT_RF	0.09***	0.11***	−0.27***	−0.13***	−0.30***	−0.22***	−0.26***	0.00	0.29***	−0.09***	0.19***	0.20***

This table reports pairwise correlations between state variable innovations and tradeable risk factors. Panel A reports correlations among ETF premium innovations and constructed state variable innovations. Panel B reports correlations among tradeable risk factors and controls. All variables are defined in Table 1.

* Indicates statistical significance at 10% level.

** Indicates statistical significance at 5% level.

*** Indicates statistical significance at 1% level.

expected returns than stocks that covary more negatively, ceteris paribus.

We estimate the relation between stock returns (in excess of the risk-free rate) and loadings (betas) with two empirical asset pricing methods: LAP-loadings-sorted rank portfolios and Fama and MacBeth (1973) regressions of returns on LAF loadings. Therefore (except for the loadings constructed with the equity principal component, which is constructed from the entire time-series), we estimate the loading with data available to investors at the beginning of the week.

7.2. Rank portfolio tests

For the rank portfolio method, we first estimate each stock's loadings on the LAPs by repeating the factor mimicking regressions in Section 3.3 and save the weekly loadings β_{if} from fitting Eq. (3). We form quintile-rank portfolios in which the lowest rank portfolios include the 20% of stocks with the lowest estimated loadings and the high portfolios include the 20% of stocks with the highest estimated loadings. As in Section 3.3, we then form a zero-cost portfolio (the LAF) with week $t = 0$ returns that is long the equal-weighted return of the 20% of stocks with the highest β_{if} and short the 20% of stocks with the lowest β_{if}. We repeat this process using each of the seven ETF LAPs. Per Hypothesis H_2, we expect decreasing returns as the rank of the portfolio increases. To control for correlation between the loadings on LAPs and other systematic risk factors, we risk-adjust our returns by estimating the portfolio's alpha with the CAPM, Fama and French (1993) three-factor model (*FF3*), the four factor model with Fama and French (1993) three factors and Carhart's (1997) momentum factor (*FF4C*), and models with combinations of all our controls.

Our LAF specification uses a 13-week formation period to fit Eq. (3) and calculates the equal-weighted return of each of the five quintile portfolios, as described in Section 3.3. We use equal-weighted portfolios in our primary specification as low-priced stocks are associated with higher limits of arbitrage (see, for example, Falkenstein, 1996; Keim, 1999; Diether, Lee, and Werner, 2009) and thus, equal-weighted portfolios would give more weight to stocks that are theoretically more sensitive to innovations in the arbitrage market state.[23] Our rank portfolio tests confirm our intuition; however, for comparison and robustness, we perform additional analysis on three variations of LAF. For the first variation of LAF, we maintain the formation period and value-weight the returns of the five quintiles portfolios; for the second and third variations, we reconstruct the primary specification and first variation using a 26 week formation period.

Table 4 reports the results from the test of Hypothesis H_2 using the rank portfolio method with our primary specification. In Panel A, we report the average weekly returns on each of the seven sets of rank and zero-cost portfolios formed using the loadings on the seven ETF LAPs. Panel B reports CAPM, *FF3*, and *FF4C* alphas for our LAFs and the corresponding t-statistics. In Panel A of Table 4, average portfolio returns decrease for all ETFs (with the exception of the FXE 4 and H portfolios) as the rank of the portfolio increases consistent with our premises that LAFs reflect innovations in the limited arbitrage state and stocks that covary more positively (negatively) with our factor would earn relatively lower (higher) returns over time. The average (H-L) returns for the LAFs formed using the loadings on the SPY, DIA, IWV, and the equity principal component LAPs are −0.27%, −0.28%, −0.22%, and −0.32% per week, respectively, and all returns are significant at the 1% level. The returns on the GLD and IEF LAFs are −0.21% and −0.20% per

[23] See Fama and French (2008) for a discussion of the tradeoff between overemphasis of micro-stocks using equal-weighted portfolios and domination of large stocks yielding an "unrepresentative picture" using value-weighted portfolios. We exclude stocks in our market sample with prices less than $5 to reduce concerns that our results are driven by micro-stocks.

Table 4
Equal weighted returns of factor loading sorted portfolios.

	SPY	DIA	IWV	IEF	FXE	GLD	PC
Panel A – rank portfolio returns (%)							
L	0.41***	0.37***	0.30**	0.36**	0.25	0.30	0.36**
2	0.27***	0.22**	0.25**	0.26*	0.14	0.19	0.25**
3	0.23***	0.16*	0.21**	0.21*	0.10	0.13	0.20*
4	0.17**	0.12	0.16	0.19	0.07	0.11	0.14
H	0.14	0.10	0.08	0.16	0.08	0.08	0.05
H-L	−0.27***	−0.28***	−0.22***	−0.20***	−0.17*	−0.21***	−0.32***
	(−5.41)	(−4.26)	(−3.86)	(−3.64)	(−1.96)	(−2.71)	(−5.47)
Panel B – Alphas (%) on CAPM, Fama French 3 factors, and Carhart 4 factors							
MRP	−0.27***	−0.28***	−0.22***	−0.18***	−0.16*	−0.20**	−0.32***
	(−4.28)	(−3.37)	(−3.47)	(−3.24)	(−1.92)	(−2.55)	(−4.65)
FF3	−0.27***	−0.28***	−0.23***	−0.18***	−0.16*	−0.20**	−0.33***
	(−4.47)	(−3.60)	(−3.76)	(−3.33)	(−1.94)	(−2.59)	(−4.92)
FF4C	−0.32***	−0.33***	−0.25***	−0.19***	−0.15*	−0.20***	−0.34***
	(−5.77)	(−4.62)	(−4.88)	(−3.87)	(−1.92)	(−2.86)	(−5.65)
Panel C – Alphas (%) with additional controls							
FF3 and AIV	−0.25***	−0.26***	−0.23***	−0.19***	−0.15*	−0.20**	−0.31***
	(−4.29)	(−3.46)	(−3.76)	(−3.34)	(−1.85)	(−2.56)	(−4.93)
FF3 and MILLIQ	−0.27***	−0.28***	−0.23***	−0.18***	−0.16*	−0.20**	−0.33***
	(−4.47)	(−3.58)	(−3.80)	(−3.30)	(−1.88)	(−2.54)	(−4.94)
All Controls	−0.21***	−0.24***	−0.19***	−0.14***	−0.09	−0.14**	−0.27***
	(−3.41)	(−3.59)	(−3.65)	(−3.00)	(−1.20)	(−2.16)	(−4.86)
All Controls plus SENT	−0.23***	−0.26***	−0.19***	−0.16***	−0.08	−0.12*	−0.27***
	(−3.44)	(−3.61)	(−3.58)	(−3.08)	(−1.09)	(−1.96)	(−5.05)

This table reports equal-weighted returns and alphas for factor loading portfolios sorted on covariance with ETF limited arbitrage proxy innovations (LAP). Portfolios are formed at the start of each week using the factor loadings estimated from the previous 13 weeks of returns. The zero-cost portfolios are long the highest rank portfolio (H) and short the lowest rank portfolio (L). Panel A reports the average weekly returns of five rank portfolios and the corresponding zero-cost portfolios (H-L) with the seven ETF LAPs. Panel B reports the average weekly CAPM, *FF3* (Fama and French, 1993), and *FF4C* (Carhart, 1997) alphas and corresponding t-statistics. Panel C reports the average weekly alphas from fitting regressions with the zero-cost portfolio returns as dependent variables and combinations of controlling factors as independent variables. All Factors controls for *MOM, AIV, MILLIQ, VIX*, and *STREV* and *All Factors plus Sent* additionally controls for *SENT* (Baker and Wurgler, 2006). See Table 1 for a description of the ETF LAP. T-statistics are reported in parentheses.
* Indicates statistical significance at 10% level.
** Indicates statistical significance at 5% level.
*** Indicates statistical significance at 1% level.

week and also significant at the 1% level. The returns on the FXE zero-cost portfolio is −0.17% per week and significant at the 10% level.

In Panel B, consistent with the returns of the rank portfolios, the CAPM, *FF3*, and *FF4C* alphas are also all economically and statistically significant. The SPY LAF generates alphas that range from −0.32% to −0.27% per week and significant at the 1% level. The DIA and IWV LAFs have *FF4C* alphas of −0.33% and −0.25%, respectively, and are statistically significant at the 1% level; the CAPM and *FF3* alphas are significant at the 1% level, as well. The IEF, FXE, and GLD LAFs have smaller *FF4C* alphas of −0.19%, −0.15%, and −0.20%, and are statistically significant at the 1%, 10%, and 1% level, respectively.

In Panel C, we additionally control for *AIV, MILLIQ*, mean reversion, and additional shocks (see Section 5.4) that might selectively impact either the ETF or the underlying assets. We report the alphas from models augmented with combinations of the *AIV, MILLIQ, STREV, SENT*, and *VIX* factors. For the four equity LAFs, augmenting the factor models with the *AIV* and *MILLIQ* factors reduces the alphas by an economically insignificant two to seven basis points below the alphas found with *FF4C*, and the alphas remain statistically significant at the 1% level. Augmenting the factor models with all of the controls reduces the equity and the equity principal component LAF alphas by up to an additional four basis points and, again, the alphas remain significant at the 1% level. The alphas of the SPY LAF range from −0.32% to −0.27%. For the DIA LAF and IWV LAF, the alphas range from −0.33% to −0.28% and −0.25% to −0.22%, respectively. The equity principal component LAF alphas are larger, ranging from −0.31% to −0.27%. Overall, the findings in Table 4 support our hypotheses that LAFs capture innovations in systematic limits of arbitrage, and as such, are negatively priced in the cross-section of returns (Hypothesis H_2). Additionally, the reduction in alphas from Panel B to Panel C provides support for Hypotheses H_3 and H_4, that our LAFs contain information both about known limits to arbitrage, aggregate idiosyncratic volatility and market illiquidity, and beyond those limits.

Tables 5 through 7 display the findings from repeating the analyses of Table 4 using the three alternative specifications of our LAF. In Table 5, when we value-weight the quintile portfolios, the magnitudes and significance of the SPY LAF and alphas are quantitatively similar to Table 4 with the exception that the significance on the *All Controls* alpha drops to the 5% level. However, the magnitudes of the DIA and IWV LAFs and alphas drop by two to seven basis points and the significances drop, for the most part, to the 5% level; the significances of IWV on the *All Controls* and *All Controls w/ Sent* alphas drop to the 14% and 11% levels, respectively. Further, all of the significances on the non-equity ETFs drop below the 10% level.

In Table 6, where we equal-weight the quintile portfolios and use a 26-week formation period, the magnitudes of all of the LAFs and alphas are generally lower (with the exception of GLD) compared to the findings in Table 4; however, the equity LAFs and *FF4C* alphas remain significant at the 1% level and the non-equity LAFs and *FF4C* alphas significance is quantitatively similar to Table 4. With *All Controls plus Sent*, the significance of the DIA alpha remains at the 1% level; the significance on the SPY alpha drops

Table 5
Value weighted returns of factor loading sorted portfolios.

	SPY	DIA	IWV	IEF	FXE	GLD	PC
Panel A – rank portfolio returns (%)							
L	0.33***	0.26*	0.14	0.23	0.15	0.16	0.20
2	0.23***	0.14	0.12	0.19	0.09	0.16	0.17
3	0.18**	0.11	0.07	0.15	0.09	0.13	0.05
4	0.12	0.04	0.02	0.13	0.07	0.08	−0.01
H	0.05	−0.00	−0.04	0.12	0.09	0.07	−0.10
H-L	−0.28***	−0.26**	−0.18*	−0.11	−0.06	−0.09	−0.30***
	(−3.61)	(−2.55)	(−1.95)	(−1.13)	(−0.45)	(−0.69)	(−3.23)
Panel B – Alphas (%) on CAPM, Fama French 3 factors, and Carhart 4 factors							
MRP	−0.28***	−0.26**	−0.18**	−0.08	−0.06	−0.08	−0.30***
	(−3.49)	(−2.39)	(−2.20)	(−0.90)	(−0.48)	(−0.62)	(−3.60)
FF3	−0.28***	−0.25**	−0.19**	−0.07	−0.06	−0.08	−0.32***
	(−3.59)	(−2.47)	(−2.35)	(−0.93)	(−0.48)	(−0.60)	(−3.55)
FF4C	−0.34***	−0.31***	−0.21**	−0.08	−0.05	−0.07	−0.33***
	(−4.46)	(−3.13)	(−2.57)	(−1.08)	(−0.38)	(−0.61)	(−4.11)
Panel C – Alphas (%) with additional controls							
FF3 and AIV	−0.26***	−0.23**	−0.19**	−0.07	−0.04	−0.06	−0.29***
	(−3.48)	(−2.27)	(−2.33)	(−0.88)	(−0.34)	(−0.46)	(−3.44)
FF3 and MILLIQ	−0.27***	−0.25**	−0.19**	−0.07	−0.04	−0.07	−0.32***
	(−3.57)	(−2.37)	(−2.33)	(−0.89)	(−0.35)	(−0.56)	(−3.74)
All Controls	−0.22**	−0.19*	−0.12	0.01	0.00	0.03	−0.22***
	(−2.36)	(−1.91)	(−1.49)	(0.11)	(0.01)	(0.30)	(−2.87)
All Controls plus SENT	−0.25***	−0.23**	−0.13	−0.02	0.02	0.07	−0.24***
	(−2.61)	(−2.22)	(−1.61)	(−0.27)	(0.20)	(0.66)	(−3.03)

This table reports average value-weighted returns and alphas for factor loading portfolios sorted on covariance with ETF limited arbitrage proxy innovations (LAP). Portfolios are formed at the start of each week using the factor loadings estimated from the previous 13 weeks of returns. The zero-cost portfolios are long the highest rank portfolio (H) and short the lowest rank portfolio (L). Panel A reports the average weekly returns of five rank portfolios and the corresponding zero-cost portfolios (H-L) with the seven ETF LAPs. Panel B reports the average weekly CAPM, *FF3* (Fama and French, 1993), and *FF4C* (Carhart, 1997) alphas and corresponding t-statistics. Panel C reports the average weekly alphas from fitting regressions with the zero-cost portfolio returns as dependent variables and combinations of controlling factors as independent variables. All Factors controls for MOM, *AIV, MILLIQ, VIX*, and *STREV* and *All Factors plus Sent* additionally controls for *SENT* (Baker and Wurgler, 2006). See Table 1 for a description of the ETF LAP. *T*-statistics are reported in parentheses.

* Indicates statistical significance at 10% level.

** Indicates statistical significance at 5% level.

*** Indicates statistical significance at 1% level.

to the 10% level; and the significance on the IWV alpha becomes insignificant at the 14% level. The principal component LAF and alphas drop between six to nine basis points with significances remaining at the 1% level. The effect of the longer formation period is more mixed for the non-equity ETFs. The LAFs and alphas of FXE and GLD generally increase with improved significance while the IEF LAF and alphas are generally reduced similar to the stock ETFs.

In Table 7, where we both value-weight the quintile portfolios and use a 26 week formation period, the SPY, principal component, and GLD LAFs and alphas remain significant and the remainder of the LAFs and alphas become insignificant.

The findings reported in Tables 5 through 7 using alternative LAF specifications support our expectation of more robust pricing information using equal-weighted quintile portfolios as smaller and lower priced stocks should be more sensitive to limited arbitrage. The findings using a 26-week LAF formation period suggests that the average stock's sensitivity to innovations in systematic limited arbitrage is time-varying. However, though increasing the formation period generally lowers the magnitudes and significances of the LAFs and alphas, SPY, DIA, GLD, and the equity principal component LAFs, for the most part, remain significant at the 1% and 5% levels in Table 6. From Table 2, these ETFs have higher trading volumes, lower Amihud price impact, and lower bid-ask spreads (SPY and DIA) than the remaining ETFs. Also, SPY, DIA, and GLD all have lower holdings complexity than IWV, and SPY and DIA have the longest time-series. Based on the above analyses, for the remainder of the priced-risk tests, we use our primary LAF specification. Further, we omit the non-equity ETFs, as our test assets are equities, the equity ETFs likely more precisely capture the limits to arbitrage that specifically impact equity markets, and our equity ETFs have longer time-series.

7.3. Fama–MacBeth regressions

To test Hypotheses H_3 and H_4 and provide additional evidence for Hypothesis H_2, we perform Fama and MacBeth (1973) regressions, controlling for several factors. For the Fama–MacBeth method, we use test portfolios to reduce the noise in returns and portfolio formation and include control factors that maintain significant variation in LAP loadings. At the beginning of each week, we form two sets of 25 value-weighted test portfolios. For the first set, we sort stocks into 25 portfolios based on their prior 13-week loadings on the LAPs (25 LAF). Sorting portfolios by LAF loadings allows for both maximal dispersion across the factor loadings and reduction in the errors-in-variable problem. For the second set, we obtain 25 portfolios each week by first sorting stocks into quintiles based on the stock's market capitalization four weeks prior to the current week and then sorting each of those quintiles into quintiles based on the stock's price four weeks prior to the current week. Size and price are arguably related to limits to arbitrage; thus sorting by these characteristics should generate cross-sectional variation in sensitivities to systematic limited arbitrage. For example, compared to large and high-priced stocks, small and low-priced stocks garner less attention, are more difficult to short and, typically, are not listed on option exchanges (Falkenstein, 1996; Keim, 1999; Diether, Lee, and Werner, 2009). Arbitrageurs and informed investors would face greater thresholds to arbitrage mispricing with such stocks.

Table 6
Equal weighted returns of factor loading sorted portfolios with 26 week formation period.

	SPY	DIA	IWV	IEF	FXE	GLD	PC
Panel A – rank portfolio returns (%)							
L	0.34***	0.34***	0.29**	0.32**	0.28	0.35	0.35**
2	0.26***	0.22**	0.25**	0.24*	0.16	0.22	0.25**
3	0.23***	0.16*	0.22**	0.21*	0.12	0.13	0.23**
4	0.18**	0.14	0.18	0.20	0.10	0.09	0.16
H	0.17*	0.14	0.14	0.18	0.06	0.05	0.09
H-L	−0.17***	−0.20***	−0.14***	−0.14**	−0.22**	−0.30***	−0.25***
	(−3.22)	(−2.86)	(−2.71)	(−2.56)	(−2.40)	(−3.92)	(−4.42)
Panel B – Alphas (%) on CAPM, Fama French 3 factors, and Carhart 4 factors							
MRP	−0.18***	−0.20**	−0.14**	−0.13*	−0.21**	−0.29***	−0.25***
	(−2.69)	(−2.33)	(−2.34)	(−1.95)	(−2.09)	(−3.40)	(−3.55)
FF3	−0.18***	−0.21**	−0.15**	−0.13*	−0.20**	−0.29***	−0.27***
	(−2.79)	(−2.56)	(−2.43)	(−1.96)	(−2.04)	(−3.34)	(−3.70)
FF4C	−0.23***	−0.26***	−0.16***	−0.13**	−0.19**	−0.28***	−0.28***
	(−3.88)	(−3.59)	(−2.95)	(−2.40)	(−2.14)	(−4.13)	(−4.16)
Panel C – Alphas (%) with additional controls							
FF3 and AIV	−0.15**	−0.18**	−0.15**	−0.13**	−0.20**	−0.30***	−0.26***
	(−2.39)	(−2.31)	(−2.41)	(−2.07)	(−2.01)	(−3.51)	(−3.61)
FF3 and MILLIQ	−0.17***	−0.20**	−0.15**	−0.12*	−0.20**	−0.28***	−0.26***
	(−2.69)	(−2.53)	(−2.38)	(−1.91)	(−2.03)	(−3.40)	(−3.67)
All Controls	−0.09	−0.13**	−0.06	−0.08*	−0.13	−0.23***	−0.18***
	(−1.64)	(−2.21)	(−1.38)	(−1.68)	(−1.60)	(−4.10)	(−3.19)
All Controls plus Sent	−0.11*	−0.17***	−0.07	−0.12**	−0.14*	−0.23***	−0.19***
	(−1.94)	(−2.76)	(−1.47)	(−2.49)	(−1.69)	(−3.85)	(−3.26)

This table reports equal-weighted returns and alphas for factor loading portfolios sorted on covariance with ETF limited arbitrage proxy innovations (LAP). Portfolios are formed at the start of each week using the factor loadings estimated from the previous 26 weeks of returns. The zero-cost portfolios are long the highest rank portfolio (H) and short the lowest rank portfolio (L). Panel A reports the average weekly returns of five rank portfolios and the corresponding zero-cost portfolios (H-L) with the seven ETF LAPs. Panel B reports the average weekly CAPM, *FF3* (Fama and French, 1993), and *FF4C* (Carhart, 1997) alphas and corresponding t-statistics. Panel C reports the average weekly alphas from fitting regressions with the zero-cost portfolio returns as dependent variables and combinations of controlling factors as independent variables. All Factors controls for MOM, *AIV, MILLIQ, VIX*, and *STREV* and *All Factors plus Sent* additionally controls for *SENT* (Baker and Wurgler, 2006). See Table 1 for a description of the ETF LAP. *T*-statistics are reported in parentheses.

* Indicates statistical significance at 10% level.

** Indicates statistical significance at 5% level.

*** Indicates statistical significance at 1% level.

We estimate the price of arbitrage risk using the two-stage Fama–MacBeth procedure. In the first stage, at the beginning of each calendar year, we regress each of the 25 test portfolio's weekly prior calendar year returns on the prior calendar year weekly LAF of interest and control factors. We capture these 25 estimated betas (quantity of risk) and use them as regressands in the second-stage regressions as follows. In the second stage, for each weekly cross-section in each calendar year, we regress the 25 test portfolio's returns on their captured LAF betas estimated in the first-stage to estimate the gammas (the prices of risk) and their corresponding test-statistics. Table 8 reports these parameter estimates, i.e. the price of risk, and associated t-statistics from the second-stage of the procedure with the SPY LAF; Table 9 reports estimates with the DIA, IWV, and the equity principal component LAFs, respectively.

To begin, we estimate the price of arbitrage risk using the betas estimated from the first-stage regression with SPY LAF. The first five columns of Table 8 report the coefficients using the 25 LAF portfolios. The last five columns report the coefficients using the 25 portfolios sorted by size and price. In Table 8, Columns (1) and (6), we include the SPY LAF and *MRP* as independent variables in the regressions.[24] In Columns (2) and (7), we add the factor loadings on *SMB, HML*, and *MOM* as additional independent variables with LAF and *MRP*. Finding negative and significant coefficients on SPY LAF in the first and second columns of each set provides further evidence that stocks with more positive loadings on LAF have lower expected returns than stocks with more negative loadings and that LAF is negatively priced in the cross-section of stock returns (Hypothesis H_2). In Columns (3) and (8), we additionally include the factor loadings on the known limits to arbitrage risk factors, *AIV* and *MILLIQ*. Finding that the coefficients on SPY LAF remain negative when controlling for these well-known limits of arbitrage, provides evidence that the LAF contains information about systematic limited arbitrage beyond individual limits to arbitrage (Hypothesis H_4). Although we posit that each well-known limit will systematically limit arbitrage and that channel of impact would be negatively priced; we do not predict the sign of the coefficient on that limit as the limit may also impact cross-sectional stock pricing through other channels. In Columns (4) and (9), we exclude the LAF. Finally, in Columns (5) and (10), we include the loading on *SENT* in our cross-sectional regressions. We include *SENT* along with *VIX, STREV*, and the *FF4C*, to control for systematic shocks that might selectively impact the ETF price or the price of the underlying asset. As such, we control for systematic shocks to the ETF premium unrelated to innovations in the ability to perform arbitrage.

The price of risk estimates support that LAF is priced in the cross-section of returns. Consistent with the rank portfolio results and Hypothesis H_2, we find the price of arbitrage risk is negative and significant in all ten specifications in Table 8.[25] In the first set

[24] We include *MRP* in all of our Fama–MacBeth regressions because we included it in Eq. (5) to estimate the factor loadings.

[25] As our premise is that LAFs capture innovations in limited arbitrage, the price of arbitrage risk is the coefficient on LAF in the second-stage regression as it reflects the average relation between test portfolio returns and the quantity (the first stage betas) of test portfolio risk.

Table 7
Value weighted returns of factor loading sorted portfolios with 26 week formation period.

	SPY	DIA	IWV	IEF	FXE	GLD	PC
Panel A – rank portfolio returns (%)							
L	0.30***	0.21	0.12	0.18	0.17	0.33	0.22
2	0.23***	0.16	0.15	0.22*	0.14	0.21	0.12
3	0.19**	0.09	0.07	0.16	0.15	0.09	0.10
4	0.08	0.02	0.03	0.15	0.04	0.03	0.00
H	0.08	0.04	0.04	0.12	−0.04	−0.00	−0.02
H-L	−0.22***	−0.17	−0.08	−0.06	−0.22	−0.33***	−0.24***
	(−2.64)	(−1.64)	(−0.91)	(−0.64)	(−1.41)	(−2.64)	(−2.69)
Panel B – Alphas (%) on CAPM, Fama French 3 factors, and Carhart 4 factors							
MRP	−0.23***	−0.17	−0.08	−0.03	−0.22	−0.32**	−0.24***
	(−2.74)	(−1.63)	(−0.95)	(−0.36)	(−1.51)	(−2.50)	(−2.59)
FF3	−0.24***	−0.17*	−0.10	−0.04	−0.19	−0.31**	−0.28***
	(−2.88)	(−1.72)	(−1.22)	(−0.45)	(−1.24)	(−2.54)	(−2.94)
FF4C	−0.31***	−0.24**	−0.12	−0.05	−0.17	−0.30***	−0.30***
	(−4.00)	(−2.56)	(−1.50)	(−0.62)	(−1.18)	(−3.19)	(−3.23)
Panel C – Alphas (%) with additional controls							
FF3 and AIV	−0.19**	−0.14	−0.09	−0.04	−0.21	−0.34***	−0.26***
	(−2.45)	(−1.45)	(−1.17)	(−0.52)	(−1.35)	(−2.88)	(−2.81)
FF3 and MILLIQ	−0.22***	−0.17*	−0.10	−0.03	−0.18	−0.30**	−0.27***
	(−2.71)	(−1.69)	(−1.26)	(−0.38)	(−1.22)	(−2.53)	(−2.81)
All Controls	−0.11	−0.07	0.04	0.03	−0.12	−0.23***	−0.15*
	(−1.38)	(−0.84)	(0.56)	(0.44)	(−0.81)	(−2.92)	(−1.72)
All Controls plus SENT	−0.16**	−0.13	0.02	−0.03	−0.14	−0.22***	−0.17*
	(−1.97)	(−1.48)	(0.19)	(−0.46)	(−0.95)	(−2.66)	(−1.96)

This table reports value-weighted returns and alphas for factor loading portfolios sorted on covariance with ETF limited arbitrage proxy innovations (LAP). Portfolios are formed at the start of each week using the factor loadings estimated from the previous 26 weeks of returns. The zero-cost portfolios are long the highest rank portfolio (H) and short the lowest rank portfolio (L). Panel A reports the average weekly returns of five rank portfolios and the corresponding zero-cost portfolios (H-L) with the seven ETF LAPs. Panel B reports the average weekly CAPM, *FF3* (Fama and French, 1993), and *FF4C* (Carhart, 1997) alphas and corresponding t-statistics. Panel C reports the average weekly alphas from fitting regressions with the zero-cost portfolio returns as dependent variables and combinations of controlling factors as independent variables. All Factors controls for MOM, *AIV, MILLIQ, VIX*, and *STREV* and *All Factors plus Sent* additionally controls for *SENT* (Baker and Wurgler, 2006). See Table 1 for a description of the ETF LAP. T-statistics are reported in parentheses.
* Indicates statistical significance at 10% level.
** Indicates statistical significance at 5% level.
*** Indicates statistical significance at 1% level.

of regressions (25 LAF portfolios), the point estimates of the price of arbitrage risk range from −0.218% to −0.140% with *t*-statistics of the average second-stage regression coefficients that range from −2.47 to −3.61. Adding the *FF4C* in Specification (2) does not impact the level of statistical significance of the coefficients. When adding *AIV, MILLIQ, STREV*, and *VIX*, the significance falls somewhat, but is still significant at the 5% level. When using the 25 size and price-sorted portfolios as test assets, the price of risk estimates are more negative and statistically significant. The t-statistics of the four specifications that include the first-stage SPY LAF betas range from (3.31) to (5.14). Adding the other factor betas to the model does not impact the level of significance of our SPY LAF beta. These price of arbitrage risk estimates are consistent with our rank and zero-cost portfolio results and support our Hypothesis H_2. Further, as the price of risk in Specifications (3) and (8) remains negative and significant with the addition of *AIV* and *MILLIQ*, our findings are consistent with Hypothesis H_4, are robust to the inclusion of multiple factors, and, thus, support that LAFs contain information about systematic limited arbitrage beyond individual limits to arbitrage.

In Table 9, we replicate the price of arbitrage risk estimates from Table 8, but use LAF constructed from the premium innovations of DIA, IWV, and the equity principal component. In Panel A, we use the 25 LAF portfolios as the test assets and in Panel B we use the 25 size and price-sorted portfolios. We estimate the price of arbitrage risk with three specifications. The first specification includes the LAF and *MRP* as independent variables in the regressions; the second specification adds the factor loadings on *SMB, HML*, and *MOM* as additional independent variables with LAF and *MRP*. The third specification is includes the factor loadings on the remainder of our controls.

Consistent with the findings in Table 8 for SPY LAF, the price of risk estimates support that DIA, IWV, and the equity principal component LAFs are priced in the cross-section of returns. The price of risk for the DIA LAF ranges from −0.300% to −0.244%, and are all significant at the 1% level. The price of risk for the principal component LAF ranges from −0.199% to −0.142%, and are all statistically significant at the 10% level or better. The price of risk in the first two specifications for the IWV LAF are −0.195% and −0.194%, and are both significant at the 5% level or better, respectively. The price of risk in the specification with all controls is −0.142%, which is economically large, but marginally significant (*t*-statistic of −1.63). In Panel B, we replicate the estimation from Panel A, using the 25 portfolios sorted by size and price. The price of risk for all nine estimates (three LAFs with three specifications each) range from −0.530% to −0.283%, and are all significant at the 1% level. The FF4 and the other factors largely do not impact the size and significance of the price of arbitrage risk. This supports Hypothesis H_4, that the LAF captures information beyond *AIV* and *MILLIQ*.

7.4. Summary of portfolio and FM tests

Overall, the portfolio alphas and price of risk estimates in Tables 4 through 9 support our Hypotheses H_2, H_3, and H_4 and that SPY, DIA, IWV, and the equity principal component LAFs capture innovations in limited arbitrage. The average return and portfolio alphas of the equity LAP loading-sorted portfolios are

Table 8
Fama–MacBeth regressions with SPY LAF.

	25 LAF					5 × 5 size × price				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
LAF	−0.218***	−0.204***	−0.140**		−0.157**	−0.282***	−0.463***	−0.434***		−0.448***
	(−3.61)	(−3.42)	(−2.32)		(−2.47)	(−3.31)	(−5.14)	(−4.67)		(−4.42)
MRP	0.122	0.125	0.126	0.125	0.131	0.149	0.097	0.112	0.108	0.113
	(1.36)	(1.39)	(1.39)	(1.39)	(1.43)	(1.53)	(1.09)	(1.25)	(1.21)	(1.26)
SMB		−0.142**	−0.149**	−0.124*	−0.146**		0.031	0.059	0.057	0.073
		(−2.29)	(−2.18)	(−1.90)	(−2.00)		(0.62)	(1.16)	(1.10)	(1.37)
HML		−0.037	−0.006	−0.010	−0.014		0.138**	0.064	0.076	0.035
		(−0.66)	(−0.10)	(−0.17)	(−0.22)		(2.19)	(0.95)	(1.09)	(0.48)
MOM		−0.194**	−0.183*	−0.207**	−0.181*		−0.052	−0.016	−0.017	−0.056
		(−2.10)	(−1.85)	(−2.21)	(−1.68)		(−0.59)	(−0.17)	(−0.19)	(−0.60)
AIV			0.008	0.002	−0.004			−0.036	−0.041	0.038
			(0.08)	(0.02)	(−0.04)			(−0.30)	(−0.34)	(0.31)
MILLIQ			0.026	−0.001	0.022			0.099	0.096	0.112
			(0.26)	(−0.01)	(0.21)			(0.73)	(0.72)	(0.80)
STREV			0.112	0.116	0.118			0.603***	0.639***	0.622***
			(0.94)	(1.04)	(0.93)			(5.08)	(5.43)	(5.27)
VIX			−0.091	−0.112	−0.083			0.065	0.087	0.021
			(−0.84)	(−1.06)	(−0.73)			(0.52)	(0.70)	(0.16)
SENT					0.039					−0.037
					(0.32)					(−0.23)
Obs.	883	883	883	883	831	883	883	883	883	831

This table reports estimated coefficients from the second stage of a two-stage Fama and MacBeth (1973) procedure. The first stage fits annual time-series regressions with returns on 25 test portfolios regressed on the SPY LAF and risk factors. The regressands include combinations of LAF, *FF3*, and *MOM, AIV, MILLIQ, VIX, STREV*, and *SENT* (described in Table 1). The second stage fits weekly cross-sectional regressions with returns on the 25 test portfolios regressed on the prior 52 week betas from the first stage. Coefficient estimates from the second-stage and corresponding t-statistics are reported. The first set of columns reports coefficients using portfolios formed by sorting stocks into 25 portfolios based on their loadings on SPY LAF. The second set of columns reports coefficients using portfolios formed by first sorting stocks into quintiles based on the stock's market capitalization four weeks prior to the current week and then sorting each of those quintiles into quintiles based on the stock's price. *T*-statistics are reported in parentheses.

* Indicates statistical significance at 10% level.
** Indicates statistical significance at 5% level.
*** Indicates statistical significance at 1% level.

negative and significant, and, similarly, the price of arbitrage risk for each equity LAF, as estimated on the LAF betas using the Fama and MacBeth (1973) procedure, are also consistently negative and significant. Further, our finding in Tables 8 and 9 that the significance of coefficients on LAFs is robust to controls for *AIV* and *MILLIQ* provides additional support that LAFs contain information beyond individual limits to arbitrage.

8. Supplemental analysis and robustness

In this section, we perform supplemental analyses and additional robustness tests. In Section 8.1, we perform supplemental analyses using the LAP-loadings rank portfolio method to provide further support for Hypotheses H_3 and H_4: LAFs are relevant measures of limited arbitrage after controlling for information in the *AIV* and *MILLIQ* limits to arbitrage and provide information about systematic limited arbitrage beyond any specific limit. In Section 8.2, we report and analyze characteristics of the H and L quintiles of our LAP-loadings sorted portfolios to test the implication of our hypothesis that the L portfolio would contain stocks with characteristics that suggest relatively lower returns and the H portfolio would contain stocks with characteristics that suggest relatively higher returns as systematic arbitrage became more limited. In Section 8.3, we report the results from three additional robustness tests. First, we replicate our portfolio results outside of the financial crisis period. Second, we winsorize our LAPs prior to constructing LAFs. Third, we report additional Fama and MacBeth (1973) price of arbitrage risk estimates using individual stocks rather than test portfolios.

8.1. Supplemental analysis with the rank portfolio method

We perform supplemental analyses using the rank portfolio method to provide further support for Hypotheses H_3 and H_4: that LAFs are relevant measures of limited arbitrage after controlling for information in the *AIV* and *MILLIQ* limits to arbitrage and capture information about systematic limited arbitrage beyond any specific limit.

To perform this supplemental test, we repeat the portfolio tests performed in Section 7.2 (Table 4) with an additional sequential sort in the rank portfolio construction that allows the portfolios to maintain dispersion in the LAP loadings while removing dispersion in the *AIV* and *MILLIQ* loadings. At the beginning of each week, we first sort stocks into quintiles based on their loading on the *AIV* and *MILLIQ* state variable innovations. Then, within each of these quintiles, as in Section 7.2, we form rank quintile portfolios with loadings on the ETF LAP. Additionally, we then form an average portfolio for each ETF LAP-loadings rank (L through H) that spans the *AIV* and *MILLIQ* quintiles. Since these average portfolios contain equal parts of each of the *AIV* and *MILLIQ* quintiles, dispersion across *AIV* and *MILLIQ* should be eliminated. Finding that the rank LAP-loadings portfolio returns decrease monotonically both within the *AIV* and *MILLIQ* quintiles and the spanned average would provide additional evidence that the LAPs and LAFs have information about systematically limited arbitrage beyond the specific limits.

Table 10 reports the returns of portfolios formed from the sequential sorts using the SPY LAP. In the first (second) set of returns in each panel, we remove the dispersion in *AIV* (*MILLIQ*) loadings. The column labeled *Avg* reports the returns of the portfolios spanning the specified limits' rank. In the top section (block of five rows) of each panel, we report the average returns of the rank portfolios; in the middle section, we report the average zero-cost portfolio return, *FF3* alpha, and the corresponding t-statistics. In the bottom section of each panel, we report the average LAP loadings on the stocks in the corresponding rank and zero-cost portfolios in the top and middle sections of the panel.

Table 9
Fama–MacBeth regressions with LAF.

Panel A: 25 LAF

	DIA			IWV			PC‘		
LAF	−0.300***	−0.247***	−0.244***	−0.195***	−0.194**	−0.142	−0.199***	−0.169**	−0.142*
	(−3.71)	(−3.24)	(−2.92)	(−2.61)	(−2.48)	(−1.63)	(−2.86)	(−2.52)	(−1.81)
MRP	0.019	0.024	0.020	0.070	0.086	0.089	0.077	0.088	0.090
	(0.16)	(0.21)	(0.17)	(0.60)	(0.73)	(0.73)	(0.65)	(0.74)	(0.74)
SMB		−0.136*	−0.087		−0.068	0.011		−0.089	−0.029
		(−1.71)	(−0.90)		(−0.92)	(0.11)		(−1.16)	(−0.29)
HML		0.002	0.046		−0.025	−0.031		−0.012	0.006
		(0.03)	(0.55)		(−0.41)	(−0.44)		(−0.20)	(0.08)
MOM		−0.286**	−0.298**		−0.209*	−0.181		−0.206*	−0.171
		(−2.42)	(−2.09)		(−1.85)	(−1.27)		(−1.86)	(−1.18)
AIV			0.025			−0.036			−0.127
			(0.19)			(−0.28)			(−0.96)
MILLIQ			0.106			0.166			0.203
			(0.78)			(1.30)			(1.58)
STREV			0.173			0.213*			0.204
			(1.19)			(1.68)			(1.61)
VIX			0.069			−0.006			0.026
			(0.48)			(−0.04)			(0.18)
SENT			0.097			0.231*			0.255*
			(0.64)			(1.68)			(1.79)
Obs.	624	624	572	521	521	469	521	521	469

Panel A: 25 LAF

	DIA			IWV			PC		
LAF	−0.373***	−0.530***	−0.571***	−0.283***	−0.430***	−0.430***	−0.342***	−0.436***	−0.524***
	(−3.49)	(−4.27)	(−4.00)	(−3.07)	(−4.22)	(−3.67)	(−3.40)	(−3.95)	(−3.98)
MRP	0.100	−0.003	0.008	0.097	0.055	0.064	0.113	0.056	0.063
	(0.79)	(−0.03)	(0.07)	(0.80)	(0.46)	(0.52)	(0.94)	(0.47)	(0.52)
SMB		0.092	0.147**		0.048	0.099		0.054	0.082
		(1.54)	(2.23)		(0.89)	(1.62)		(0.99)	(1.38)
HML		0.144*	0.003		0.177**	0.018		0.134*	0.002
		(1.75)	(0.03)		(2.41)	(0.23)		(1.86)	(0.02)
MOM		−0.131	−0.055		−0.001	−0.019		−0.060	−0.110
		(−1.13)	(−0.43)		(−0.01)	(−0.16)		(−0.56)	(−0.92)
AIV			0.088			0.195			0.181
			(0.53)			(1.10)			(1.03)
MILLIQ			0.088			0.200			0.016
			(0.47)			(1.06)			(0.09)
STREV			0.725***			0.556***			0.462***
			(4.58)			(3.29)			(2.85)
VIX			0.044			0.257			0.094
			(0.26)			(1.52)			(0.53)
SENT			−0.272			−0.178			−0.133
			(−1.30)			(−0.88)			(−0.61)
Obs.	624	624	572	521	521	469	521	521	469

This table reports estimated coefficients from the second stage of a two-stage Fama and MacBeth (1973) procedure. The first stage fits annual time-series regressions with returns on 25 test portfolios regressed on the DIA, IWV, and PC LAF and risk factors. The regressands include combinations of LAF, *FF3*, and *MOM, AIV, MILLIQ, VIX, STREV*, and *SENT* (described in Table 1). The second stage fits weekly cross-sectional regressions with returns on the 25 test portfolios regressed on the prior 52 week betas from the first stage. Coefficient estimates from the second-stage and corresponding t-statistics are reported. Panel A reports coefficients using portfolios formed by sorting stocks into 25 portfolios based on their loadings on LAF. Panel B reports coefficients using portfolios formed by first sorting stocks into quintiles based on the stock's market capitalization four weeks prior to the current week and then sorting each of those quintiles into quintiles based on the stock's price. *T*-statistics are reported in parentheses.

* Indicates statistical significance at 10% level.
** Indicates statistical significance at 5% level.
*** Indicates statistical significance at 1% level.

The findings from our tests reported in Table 10 support our contention that the SPY LAF contains additional information for the pricing of stocks beyond aggregate idiosyncratic volatility and market illiquidity (Hypothesis H_4). When using the *AIV* factor loading as the first sorting variable, the mean returns of the LAP-loadings rank portfolios within each *AIV* loadings quintile decrease as the rank of the portfolio increases. Therefore, regardless of the loading on *AIV*, we find the negative relation between returns and LAP loadings. The average returns on and alphas of the five zero-cost portfolios formed within the five *AIV* loading portfolios are all economically significant (ranging from −0.18% to −0.29% per week) and statistically significant at the 1% level. In the *Avg* column, the LAP-loadings sorted portfolio returns decrease as the rank of the portfolio increases and the zero-cost portfolios have average returns and alphas of −0.25% and −0.24%, respectively, significant at the 1% level. When using the *MILLIQ* factor loading as the first sorting variable, the average SPY portfolio returns, again, decrease as the rank of the portfolio increases in all of the *MILLIQ* quintiles and the corresponding zero-cost portfolios have economically and statistically significant average returns and *FF3* alphas.

In the bottom section of Table 10, we report the factor loadings of the rank and zero-cost portfolios. The average H-L factor loadings reported in the last row of the table form a U-shaped pattern across both the *AIV* and *MILLIQ* quintiles. Together, the

Table 10
Returns of limited arbitrage sequentially sorted factor loading sorted portfolios for SPY.

	AIV quintiles						*MILLIQ* quintiles					
	L	*2*	*3*	*4*	*H*	*Avg*	*L*	*2*	*3*	*4*	*H*	*Avg*
Rank portfolio average returns (%):												
L	0.50***	0.44***	0.37***	0.38***	0.32***	0.40***	0.40***	0.37***	0.38***	0.43***	0.45***	0.41***
2	0.32***	0.28***	0.24***	0.23***	0.22**	0.26***	0.26***	0.27***	0.27***	0.25***	0.31***	0.27***
3	0.26***	0.28***	0.24***	0.21***	0.17*	0.23***	0.20**	0.24***	0.22***	0.24***	0.23***	0.23***
4	0.19**	0.21***	0.19***	0.15**	0.15	0.18**	0.15	0.18**	0.19***	0.19***	0.16*	0.17**
H	0.21*	0.17*	0.18**	0.15*	0.07	0.16*	0.14	0.15*	0.16*	0.12	0.18	0.15
Zero-cost portfolio returns (%):												
H-L	−0.29***	−0.27***	−0.19***	−0.23***	−0.26***	−0.25***	−0.27***	−0.22***	−0.22***	−0.31***	−0.27***	−0.26***
	(−4.58)	(−5.49)	(−4.48)	(−4.83)	(−4.00)	(−5.39)	(−4.15)	(−4.48)	(−4.82)	(−6.51)	(−4.21)	(−5.56)
Alpha	−0.27***	−0.25***	−0.18***	−0.22***	−0.25***	−0.24***	−0.26***	−0.21***	−0.21***	−0.30***	−0.26***	−0.25***
	(−3.77)	(−4.71)	(−3.90)	(−4.01)	(−3.35)	(−4.35)	(−3.62)	(−3.79)	(−3.88)	(−5.42)	(−3.58)	(−4.39)
Factor loadings:												
L	−12.35***	−9.26***	−8.76***	−9.54***	−12.59***	−10.49***	−12.46***	−9.47***	−8.88***	−9.52***	−12.50***	−10.56***
2	−3.65***	−2.78***	−2.62***	−2.97***	−3.89***	−3.18***	−3.69***	−2.83***	−2.68***	−2.96***	−3.86***	−3.20***
3	0.52***	0.20***	0.11***	0.08	0.33***	0.25***	0.55***	0.22***	0.08**	0.06	0.32***	0.25***
4	4.81***	3.21***	2.87***	3.16***	4.67***	3.74***	4.90***	3.29***	2.88***	3.11***	4.60***	3.76***
H	14.59***	10.03***	9.34***	10.14***	14.28***	11.67***	14.86***	10.27***	9.42***	10.03***	14.22***	11.76***
H-L	26.94***	19.29***	18.10***	19.68***	26.86***	22.17***	27.33***	19.74***	18.30***	19.55***	26.72***	22.32***
	(39.22)	(39.69)	(38.80)	(38.62)	(40.66)	(39.65)	(39.48)	(39.96)	(39.02)	(39.54)	(40.46)	(39.90)

This table reports average returns and *FF3* alphas for factor loading sequentially-sorted portfolios. To remove dispersion across known limits of arbitrage, stocks are first sorted into quintiles based on their factor loadings on *AIV* and *MILLIQ* innovations and then sorted within each of these quintiles into rank portfolios based on their factor loadings on the SPY limited arbitrage proxy. In the first set of returns, the dispersion in *AIV* loadings is removed; in the second set, the dispersion in *MILLIQ* loadings is removed. The column labeled Avg reports the returns of the average portfolios spanning the *AIV* and *MILLIQ* quintiles, respectively. The top section of each set of returns reports the average returns of the rank portfolios. The middle section reports the average return, *FF3* alpha, and the corresponding t-statistics for the five zero-cost and spanned average portfolios, and the bottom section reports the loadings corresponding to the rank and zero-cost portfolios in the top and middle sections of the panels. The state variables and *FF3* alphas are described in Tables 1 and 4 respectively. *T*-statistics are reported in parentheses.

* Indicates statistical significance at 10% level.
** Indicates statistical significance at 5% level.
*** Indicates statistical significance at 1% level.

dispersion in LAP factor loadings within each *AIV* and *MILLIQ* quintile (columns of the table section) coupled with the U-shaped pattern across the *AIV* and *MILLIQ* quintiles (rows of the table section) supports that the dispersion in LAP betas is driven by factors other than *AIV* and *MILLIQ*, consistent with our Hypothesis H_4 that LAFs carry information beyond *AIV* and *MILLIQ*.

Table 11 repeats the analysis reported in Table 10, except SPY is replaced with DIA, IWV, and the equity principal component. For brevity, we only report the results from the *Avg* portfolios. The findings remain consistent with Table 10: for each DIA, IWV, and the equity principal component, returns decrease as the rank of the portfolio increases and the average returns and alphas for the zero-cost portfolios are all economically and statistically significantly negative at the 1% level.

8.2. Characteristics of LAP-loadings sorted quintile portfolios

Next, we examine the average characteristics of the stocks sorted into our H and L rank portfolios by LAP-loadings. Our hypotheses posit and our pricing tests support that, over time, holding a zero-cost portfolio, long stocks that have relatively higher returns and short stocks that have relatively lower returns when arbitrage is worsening, earns negative returns. By construction, the L portfolio holds stocks that covary more negatively with LAPs and, thus, holds stocks that perform relatively worse when arbitrage is increasing. Therefore, we expect the holdings of the L portfolio to have characteristics, on average, associated with higher limits to arbitrage.

The first two characteristics we investigate are price and market capitalization (size), as low-priced, small size stocks are identified by Falkenstein (1996), Keim (1999), and Diether, Lee, and Werner (2009) as difficult to arbitrage. In Table 12, we examine average characteristics of the stocks in the H and L quintile portfolios formed for our rank portfolio tests reported in Table 4. Panel A reports the mean price of the stocks in the H and L portfolios, the differences in the mean price, and a *t*-test of the differences. The average price of the stocks in the H portfolio is significantly greater than the L portfolio across all ETF LAPs (both equity-based and commodity-based); the difference ranges from \$7.71 (SPY) to \$18.89 (FXE). Panel B reports the means and difference with size; the average market value of the stocks in the H portfolio is generally greater than the L portfolio with varying significance. The differences for the SPY and DIA are \$1.35b and \$1.80b, respectively, with significance at the 5% level. Overall, the findings with price and size are consistent, particularly for SPY and DIA, with our contention that premium innovations (LAPs) capture innovations in limited arbitrage as sorting on LAP loadings places difficult to arbitrage, low-priced and small stocks, in the L portfolio that performs relatively poorly as premium innovations become more positive.

Panel C and D report the means and differences using non-aggregated, individual stock measures of Amihud (2002) illiquidity and Goyal and Santa-Clara (2003) idiosyncratic volatility as characteristics. We expect our LAP-loadings sorts to place more illiquid stocks with higher idiosyncratic volatility in the L portfolio, as stocks with these characteristics would have the lowest returns as systematic limited arbitrage increases. The findings in Panels C and D are consistent with our expectations. For the SPY LAP portfolios, illiquidity is about 30% lower (20.80 to 14.70) for the H portfolio. The remaining ETFs have similar decreases from the H to L portfolio, and the differences are all significant at the 1% level.

Overall, the findings in Table 12 suggest a direct connection between specific known measures of limits of arbitrage and LAF and, thereby, support the premise that the construction of our LAPs and LAFs using ETF premium innovations captures innovations in limited arbitrage. In general, when systematic limited arbitrage is worsening (the ETF LAP is increasing) the L portfolio is

Table 11
Average spanned returns of limited arbitrage sequentially sorted factor loading sorted portfolios.

Rank portfolio average returns (%):	*AIV*			MILLIQ		
	DIA	*IWV*	*PC*	*DIA*	*IWV*	*PC*
L	0.36***	0.31**	0.35**	0.36***	0.29**	0.35**
2	0.21**	0.23**	0.24**	0.21**	0.25**	0.25**
3	0.16*	0.20*	0.19*	0.16*	0.21*	0.19*
4	0.13	0.16	0.15	0.12	0.15	0.14
H	−0.25***	−0.22***	−0.28***	−0.26***	−0.20***	−0.29***
Zero-cost portfolio returns (%):						
H-L	−0.25***	−0.22***	−0.28***	−0.26***	−0.20***	−0.29***
	(−4.14)	(−4.25)	(−5.19)	(−4.34)	(−3.80)	(−5.11)
Alpha	−0.24***	−0.23***	−0.28***	−0.26***	−0.21***	−0.29***
	(−3.47)	(−4.04)	(−4.64)	(−3.64)	(−3.66)	(−4.58)
Factor loadings:						
L	−11.09***	−12.27***	−6.97***	−11.17***	−12.35***	−7.03***
2	−3.33***	−3.56***	−2.02***	−3.34***	−3.56***	−2.02***
3	0.37***	0.58***	0.33***	0.37***	0.58***	0.34***
4	4.14***	4.74***	2.72***	4.17***	4.76***	2.73***
H	12.72***	14.12***	8.02***	12.78***	14.17***	8.07***
H-L	23.81***	26.39***	14.99***	23.95***	26.52***	15.10***
	(39.71)	(40.11)	(36.62)	(40.68)	(40.00)	(37.00)

This table reports average returns and *FF3* alphas for factor loading sequentially-sorted portfolios. To remove dispersion across known limits of arbitrage, stocks are first sorted into quintiles based on their factor loading on *AIV* and *MILLIQ* innovations and then sorted within each of these quintiles into rank portfolios based on their factor loadings on the ETF limited arbitrage proxies (LAP). Each set of three columns in the table reports findings for the DIA, IWV, and *PC* LAP. In the first set of columns, the dispersion in *AIV* loadings is removed and in the second set, the dispersion in *MILLIQ* loadings is removed. The columns report the returns of the average portfolios spanning the *AIV* and *MILLIQ* quintiles, respectively. The top section of each set of columns reports the average returns of the rank portfolios. The middle section reports the average return, *FF3* alpha, and the corresponding t-statistics for the average portfolios, and the bottom section reports the average loadings corresponding to the rank and zero-cost portfolios in the top and middle sections of the panels. The state variables and *FF3* alphas are described in Tables 1 and 4, respectively. *T*-statistics are reported in parentheses.

* Indicates statistical significance at 10% level.

** Indicates statistical significance at 5% level.

*** Indicates statistical significance at 1% level.

Table 12
Characteristics of factor loading sorted portfolios.

	SPY	DIA	IWV	IEF	FXE	GLD	PC
Panel A – average rank portfolio price							
L	28.43***	29.23***	30.74***	29.18***	26.40***	39.96***	31.83***
H	36.14***	40.30***	37.38***	43.85***	45.30***	50.32***	39.96***
H-L	7.71***	11.06***	6.64***	14.68***	18.89***	10.36**	8.13***
	(4.71)	(5.04)	(2.66)	(4.87)	(4.70)	(2.01)	(3.03)
Panel B – average rank portfolio size							
L	24.70***	30.06***	30.96***	27.65***	27.69***	29.13***	31.25***
H	26.05***	31.85***	31.00***	33.03***	35.51***	34.17***	30.34***
H-L	1.35**	1.80**	0.04	5.38***	7.82***	5.04***	−0.91
	(1.99)	(2.03)	(0.04)	(5.39)	(6.03)	(4.04)	(−0.92)
Panel C – average rank portfolio illiquidity							
L	20.80***	21.17***	26.51***	25.79***	37.29***	33.26***	24.48***
H	14.70***	14.64***	17.24***	19.47***	24.78***	25.28***	14.75***
H-L	−6.10***	−6.53***	−9.27***	−6.32**	−12.51***	−7.98**	−9.73***
	(−4.87)	(−4.00)	(−3.69)	(−2.54)	(−3.48)	(−2.20)	(−4.96)
Panel D – average rank portfolio idiosyncratic volatility							
L	11.95***	11.87***	10.56***	8.87***	10.40***	9.84***	10.60***
H	10.99***	11.38***	9.37***	8.03***	9.28***	8.71***	9.36***
H-L	−0.96***	−0.49**	−1.19***	−0.85***	−1.12***	−1.13***	−1.24***
	(−5.93)	(−2.19)	(−6.67)	(−5.42)	(−4.31)	(−4.91)	(−7.50)

This table reports average characteristics for the factor loading portfolios sorted on ETF limited arbitrage proxy innovations (LAP) in Table 4. Panel A reports average prices; Panel B reports average size; Panel C reports average illiquidity; and Panel D reports average idiosyncratic volatility. The columns of each Panel report the average of the weekly characteristic for all stocks in the H and L portfolios. T-statistics are reported in parentheses.

* Indicates statistical significance at 10% level.

** Indicates statistical significance at 5% level.

*** Indicates statistical significance at 1% level.

Table 13
Fama–MacBeth regressions on individual stocks.

	(1)	(2)	(3)	(4)	(5)	(6)
LAF	−0.068***	−0.072***	−0.072***		−0.076***	−0.054***
	(−2.62)	(−3.76)	(−3.79)		(−3.71)	(−2.83)
MRP	0.188**	0.160**	0.165**	0.153**	0.176**	−0.005
	(2.25)	(2.11)	(2.16)	(2.10)	(2.28)	(−0.09)
SMB		0.025	0.025	0.023	0.028	0.041
		(0.77)	(0.77)	(0.72)	(0.82)	(1.28)
HML		0.046*	0.045*	0.042*	0.045*	0.041**
		(1.87)	(1.83)	(1.79)	(1.85)	(2.00)
MOM		−0.090*	−0.087*	−0.089**	−0.091*	−0.080*
		(−1.94)	(−1.90)	(−2.05)	(−1.89)	(−1.87)
STREV			0.087***	0.073**	0.087**	0.062**
			(2.62)	(2.32)	(2.49)	(2.13)
AIV			−0.002	−0.005	−0.001	−0.009
			(−0.09)	(−0.22)	(−0.03)	(−0.41)
Milliq			0.031	0.026	0.028	0.031
			(1.17)	(1.04)	(1.01)	(1.15)
VIX			0.025	0.021	0.026	0.019
			(0.93)	(0.84)	(0.95)	(0.75)
Sent					0.036	0.011
					(0.89)	(0.31)
Near style						−8.337
						(−0.40)
Distant style						−6.533
						(−0.31)
Obs.	885	885	885	938	833	833

This table reports estimated coefficients from the second stage of a two-stage Fama and MacBeth (1973) procedure. The first stage fits calendar year annual time-series regressions with returns on all stocks in our combined CRSP/Compustat dataset regressed on the SPY risk factor. The regressands include combinations of LAF, *FF3, MOM, AIV, MILLIQ, VIX, STREV*, and *SENT* (described in Table 1). The second stage fits weekly cross-sectional regressions with returns on all stocks regressed on the prior year conditional betas from the first stage. The columns of the table report the coefficient estimates from the second-stage regressions and corresponding t-statistics. Model (6) includes additional controls for large and small market capitalization styles. Large Cap Style is the prior 13 week return on portfolios formed from the 30% of stocks with the highest market value in our combined sample; Small Cap Style is the prior 13 week return on portfolios formed from the 30% of stocks with the lowest market value. For stocks with market values greater than the median market value of the combined sample, NEAR STYLE is set to the prior returns from the Large Cap Style portfolio and *DISTANT STYLE* is set to the prior returns from the Small Cap Style portfolio. Correspondingly, for stocks with market values less than the median market value of the combined sample, *NEAR STYLE* is set to the prior returns from the Small Cap Style portfolio and *DISTANT STYLE* is set to the prior returns from the Large Cap Style portfolio. *T*-statistics are reported in parentheses.
* Indicates statistical significance at 10% level.
** Indicates statistical significance at 5% level.
*** Indicates statistical significance at 1% level.

undesirable since it has relatively lower returns and holds low-priced,
low-liquidity, and high idiosyncratic volatility stocks. In contrast, the H portfolio is desirable since it has relatively higher returns and holds higher-priced, less illiquid, and lower idiosyncratic volatility stocks.

8.3. Additional robustness

Last, we perform three additional robustness tests. First, we consider the Financial Crisis of 2008. As our findings may be driven by the crisis and the extreme limits to arbitrage present during that period, in untabulated findings, we repeat our analyses performed in Table 4 (rank portfolio returns) and Table 8 (Fama–Macbeth regressions) with LAPs and LAFs formed on SPY, excluding the period 1 July 2008 through 30 June 2009. Similar to the returns reported in Table 4, the LAF portfolio constructed with SPY generates economically significant annualized average weekly returns of −13.0% and *FF3* alphas of −11.6%, and the returns and alphas are statistically significant at the 1% level. Likewise, with few exceptions, the coefficients estimated from the Fama–MacBeth regressions are consistent with the coefficients originally reported in Table 8.[26] Second, we consider whether our findings are driven by outliers. We winsorize our SPY LAP at 1% in each tail prior to forming the SPY LAF and, in untabulated findings, repeat our analyses performed in Table 4 and Table 8. Our findings are qualitatively and quantitatively similar to the returns reported in Tables 4 and 8. Next, we repeat the five Fama–MacBeth specifications in Table 8 using individual stocks rather than test portfolios and additionally control for near and distant market capitalization styles (Broman, 2015). We report the estimates of the price of arbitrage risk in Table 13. The price of risk estimates remain statistically significant at the 1% level (ranging from 0.00 to 0.01).

Overall, these additional robustness checks suggest that limited arbitrage risk, as captured by LAF, is relevant during periods of extreme premium innovations as well as periods of less extreme innovations and that our negative estimates of the price of arbitrage risk are robust to methodology used in our Fama–MacBeth regressions.

9. Concluding remarks

In this study, we propose a parsimonious measure of innovations in limited arbitrage: innovations in ETF premiums. Underlying our work is the premise that the extent which arbitrage is

[26] The estimates of the price of risk of the SPY *LAF* in all specifications remain negative and significant at the 5% level with the portfolios formed on the 25 *LAF* loadings. For the estimates with the portfolios sorted on lagged size and price, the coefficients remain negative and significant at the 1% level in the full specification that includes the *LAF* and all controls; the coefficients with the specifications that include MRP and the *FF3* remain negative and significant at the 10% and 1% levels, respectively.

limited is time-varying, and, therefore, limited states are systematic: highly limited states result in widespread mispricing of assets and large swings in ETF premiums which should, as ETF sponsors incentivize zero premiums, correspond almost exclusively to the impacts of limited arbitrage. Further, consistent with higher trading costs to liquidate stock that performs poorly when the market-wide ability to arbitrage worsens, we predict that stocks that payoff relatively well when arbitrage becomes systematically limited for equity traders have overall lower average returns as investors are willing to pay a premium to hold these stocks.

We construct and test tradable limited arbitrage risk factors (LAFs) using factor-mimicking portfolios sorted into quintiles by individual stock loadings on the premium innovations of six ETFs, spanning equity and three other asset classes. The LAFs are long the portfolio with the most positive loading on premium innovations and short the portfolio with the most negative loading on the premium innovations. Consistent with a common component in ETF premium innovations, we find that LAFs are positively and significantly correlated with each other. Consistent with our pricing expectations, portfolio analyses and priced risk tests find that the LAF constructed from the SPY, our longest time series available, is negatively priced in the cross-section of stock prices. Our findings are robust to inclusion of controls for mean reversion and systematic shocks such as sentiment and market volatility.

As with many studies, we are limited in our data. First, to fully test our thesis that LAFs are parsimonious measures of innovations in systematic limited arbitrage and both capture information about and beyond other factors that actually limit arbitrage would be impractical for known limits, given the sheer number of studies that identify and construct specific limits of arbitrage, and clearly impossible for as of yet undiscovered limits. Thus, to perform our tests, we controlled for two well-known, commonly-cited, and accepted equity-market limits: idiosyncratic volatility and market illiquidity (Pontiff, 2006). Second, the time-series of ETFs is relatively short; our longest ETF, SPY, began in 1993. However, we believe this period is long enough to not introduce significant bias from average news over the period and thus, rely on our findings with SPY LAF.

Overall, our findings suggest that equity ETF LAFs are a parsimonious measure of innovations in limited arbitrage, capture specific limits to arbitrage as well as capture information about systematic limited arbitrage beyond any specific limit. Further, we find that arbitrage risk pertinent to equity markets, reflected in the SPY LAF, is negatively priced in the cross-section of stocks. Additionally, our findings suggest that the information contained in equity LAFs is not solely driven by periods of extreme mispricing or crisis.

References

Acharya, V.V., Shin, H.S., Yorulmazer, T., 2013. A theory of arbitrage capital. Review of Corporate Finance Studies 2, 62–97.
Adrian, T., Rosenburgh, J., 2008. Stock returns and volatility: pricing the short-run and long-run components of market risk. Journal of Finance 63, 2997–3030.
Ali, A., Hwang, L., Trombley, M.A., 2003. Arbitrage risk and the book-to-market anomaly. Journal of Financial Economics 69, 355–373.
Ang, A., Hodrick, R., Xing, Y., Zhang, X., 2006. The cross section of volatility and expected returns. Journal of Finance 51, 259–299.
Ang, A., Papanikolaou, D., Westerfield, M.M., 2014. Portfolio choice with illiquid assets. Management Science 60, 2737–2761.
Amihud, Y., 2002. Illiquidity and stock returns: Cross-section and time-series effects. Journal of Financial Markets 5, 31–56.
Barberis, N., Thaler, R., 2003. A survey of behavioral finance. In: Constantinides, G.M., Harris, M., Stulz, R. (Eds.), Handbook of the Economics of Finance: Financial Markets and Asset Pricing. Elsevier Science B.V., Amsterdam.
Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. Journal of Finance 61, 1645–1680.
Brandt, M.W., Brav, A., Graham, J.R., Kumar, A., 2010. The idiosyncratic volatility puzzle: time trend or speculative episodes? Review of Financial Studies 23, 863–899.
Brav, A., Heaton, J.B., Li, S., 2010. The limits of the limits of arbitrage. Review of Finance 14, 157–187.
Broman, M.S., 2015. Liquidity, Style Investing, and Excess Comovement of Exchange–Traded Fund Returns Working Paper. Syracuse University.
Broman, M.S., 2016. Local Demand and Comovement Working Paper. Syracuse University.
Brunnermeier, M.K., Pedersen, L.H., 2009. Market liquidity and funding liquidity. Review of Financial Studies 22, 2201–2238.
Campbell, J.Y., Lettau, M., Malkiel, B.G., Xu, Y., 2001. Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. Journal of Finance 56, 1–43.
Cao, C., Chen, Y., Liang, B., Lo, A.W., 2013. Can hedge funds time market liquidity? Journal of Financial Economics 109, 493–516.
Carhart, M.M., 1997. On persistence in mutual fund performance. Journal of Finance 52 (1), 57–82.
Chordia, T., Roll, R., Subrahmanyam, A., 2000. Commonality in liquidity. Journal of Financial Economics 56, 3–28.
Chordia, T., Roll, R., Subrahmanyam, A., 2011. Recent trends in trading activity and market quality. Journal of Financial Economics 101, 243–263.
Chou, P., Huang, T., Yang, H., 2013. Arbitrage risk and the turnover anomaly. Journal of Banking & Finance 37, 4172–4182.
Deuskar, P., 2007. Extrapolative Expectations: Implications for Volatility and Liquidity. University of Illinois Working Paper.
Diether, K.B., Lee, K.H., Werner, I.M., 2009. It's SHO time! Short-sale price tests and market quality.. Journal of Finance 64 (1), 37–73.
Duarte, J., Kamara, A., Siegel, S., Sun, C., 2014. The Systematic Risk of Idiosyncratic Volatility Working Paper. Rice University.
Falkenstein, E.G., 1996. Preferences for stock characteristics as revealed by mutual fund portfolio holdings. Journal of Finance 51 (1), 111–135.
Fama, E.F., French, K.R., 1993. Common risk factors in the returns of stocks and bonds. Journal of Financial Economics 33, 3–56.
Fama, E.F., French, K.R., 2008. Dissecting anomalies. Journal of Finance 63, 1653–1678.
Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: empirical tests. Journal of Political Economy 81, 607–636.
Gibson, R., Mougeot, N., 2004. The pricing of systematic liquidity risk: empirical evidence from the US stock market. Journal of Banking & Finance 28, 157–178.
Goyal, A., Santa-Clara, P., 2003. Idiosyncratic risk matters!. Journal of Finance 58, 975–1007.
Herskovic, B., Kelly, B., Lustig, H., van Nieuwerburgh, S., 2014. The Common Factor in Idiosyncratic Volatility: Quantitative Asset Pricing Implications (No. w20076). National Bureau of Economic Research.
Hu, G.X., Pan, J., Wang, J., 2013. Noise as information for illiquidity. Journal of Finance 68 (6), 2341–2382.
Jegadeesh, N., 1990. Evidence of predictable behavior of security returns. Journal of Finance 45 (3), 881–898.
Keim, D.B., 1999. An analysis of mutual fund design: the case of investing in small--cap stocks. Journal of Financial Economics 51, 173–194.
Lakonishok, J., Shleifer, A., Thaler, R. and Vishny, R., 1991. Window Dressing by Pension Fund Managers (No. w3617). National Bureau of Economic Research.
Lewellen, J., Nagel, S., 2006. The conditional CAPM does not explain asset-pricing anomalies. Journal of Financial Economics 82, 289–314.
Mashruwala, C., Rajgopal, S., Shevlin, T., 2006. Why is the accrual anomaly not arbitraged away? The role of idiosyncratic risk and transaction costs. Journal of Accounting & Economics 42, 3–33.
McLean, R.D., 2010. Idiosyncratic risk, long-term reversal, and momentum. Journal of Financial and Quantitative Analysis 45, 883–906.
Merton, R.C., 1973. An intertemporal asset pricing model. Econometrica 41, 867–887.
Mitchell, M., Pulvino, T., 2012. Arbitrage crashes and the speed of capital. Journal of Financial Economics 104, 469–490.
Pasquariello, P., 2014. Financial market dislocations. Review of Financial Studies 27, 1868–1914.
Pastor, L., Stambaugh, R.F., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 642–685.
Peterson, D.R., Smedema, A.R., 2013. Idiosyncratic volatility covariance and expected stock returns. Financial Management 42, 517–536.
Pontiff, J., 2006. Costly arbitrage and the myth of idiosyncratic risk. Journal of Accounting and Economics 42 (1), 35–52.
Shleifer, A., Summers, L.H., 1990. The noise trader approach to finance. Journal of Economic Perspectives 4, 19–33.
Stambaugh, R.F., Yu, J., Yuan, Y., 2015. Arbitrage asymmetry and the idiosyncratic volatility puzzle. Journal of Finance 70 (5), 1903–1948.
Teo, M., Woo, S.-J., 2004. Style effects in the cross-section of stock returns. Journal of Financial Economics 74, 367–398.
Watanabe, A., Watanabe, M., 2008. Time-varying liquidity risk and the cross section of stock returns. Review of Financial Studies 21, 2449–2486.
Whaley, R.E., 2000. The investor fear gauge. Journal of Portfolio Management 26 (3), 12–17.