ETF Arbitrage and Return Predictability*

David C. Brown
Eller College of Management
University of Arizona
dcbrown@email.arizona.edu

Shaun William Davies
Leeds School of Business
University of Colorado, Boulder
shaun.davies@colorado.edu

Matthew C. Ringgenberg
David Eccles School of Business
University of Utah
Matthew.Ringgenberg@Eccles.Utah.edu

First Draft: July 1, 2016
This Draft: November 18, 2016
Preliminary Draft — Comments Welcome

ABSTRACT

Many finance models assume the existence of noise traders who push asset prices away from fundamental values. Yet empirically, these "animal spirits" are challenging to observe because fundamental values are inherently unobservable. We examine a novel database of trades by ETF authorized participants who specifically trade to correct violations of the law of one price. These trades allow us to measure arbitrage activity. We show that noise traders do not cancel each other out and arbitrage activity is associated with predictable price distortions. Our analysis indicates that noise traders exert a non-fundamental impact on market outcomes even when arbitrageurs are active. Thus, noise traders are not simply noise, they impact prices.

Keywords: Exchange Traded Funds, ETFs, Law of One Price, Return Predictability, Arithmetic of Active Management

JEL Classification Numbers: G12, G14

*We are grateful for the helpful comments of Jesse Blocher, Scott Cederburg, Ed Van Wesep, and conference and seminar participants at Research Affiliates, the 2016 Brevan Howard Centre for Financial Analysis' Finance Theory Group Conference, HKU, HKUST, and the University of Colorado. All errors are our own.

1 Introduction

Most, probably, of our decisions to do something positive...can only be taken as the result of animal spirits — a spontaneous urge to action rather than inaction, and not as the outcome of a weighted average of quantitative benefits multiplied by quantitative probabilities.
–John Maynard Keynes, 1936

While there is an extensive theoretical literature exploring arbitrage opportunities, empiricists have been limited in their ability to test these theories because of the many challenges to measuring and quantifying arbitrage opportunities. First, fundamental values are typically unobservable. As such, empiricists are often unable to measure mispricing. Second, the existence of arbitrageurs generates a paradox: if arbitrageurs are successful at correcting mispricings, then there may be no evidence of arbitrage opportunities in the data. Conversely, if the empiricist does observe an apparent arbitrage opportunity, it is possible that it could not actually be exploited due to limits to arbitrage.[1] Thus, while a large literature discusses the theoretical impact of noise traders and arbitrage activity (e.g., (Keynes (1936) and Shleifer and Summers (1990)), there are relatively few empirical studies which quantify noise trader behavior and the subsequent trades of arbitrageurs.

In this paper, we provide novel evidence on the indirect effects of noise traders and arbitrageurs. Specifically, we exploit the primary market mechanism for exchange traded funds (ETFs), which allows us to directly observe a daily measure of arbitrage activity on a fund-by-fund basis. Our measure of arbitrage activity is implicitly related to noise traders; such investors buy (sell) ETFs at a higher (lower) price than the price of the underlying assets. While our measure of arbitrage activity represents a lower bound on total arbitrage activity in a given ETF, we show that it has predictive power for market outcomes which are economically and statistically significant.[2]

Our results focus on the indirect effects of arbitrage activity. We document that arbitrage activity predicts future asset returns — in other words, arbitrageurs succeed at enforcing the law of one price, but they do so at the *wrong* price. Beginning with the cross-section of ETFs,

[1]See Gromb and Vayanos (2010) for a survey of the limits to arbitrage literature.

[2]While we focus specifically on the primary market for ETFs and the share creation and redemption process, it is likely that other market participants synthetically create and redeem ETF shares to engage in the same arbitrage opportunities.

measured arbitrage activity provides dramatic return predictability. A strategy which short sells ETFs that have experienced the most share creations (i.e., ETFs with the most inflows), and buys ETFs with the most share redemptions (i.e., ETFs with the most outflows), generates annual abnormal returns ranging from 5% to 13%. Turning to the time-series, ETFs experience more negative returns following inflows and more positive returns following outflows, suggesting that ETF investors collectively increase and decrease their exposure to risky assets in a systematic manner associated with investment returns. As an example, in SPY, a large ETF which accounts for almost 20% of the value in all ETFs, our analysis shows that the average investor underperformed by 66 basis points per annum. Thus, our results show that noise traders do not cancel each other out (Shleifer & Summers, 1990) and they matter for prices even in the presence of arbitrageurs.

Figure 1 shows an event time graph of the return predictability induced by trading in ETFs. We define the event date (t=0) as periods with large inflows. As money flows into ETFs, cumulative abnormal returns rise sharply (to over 1%). However, these returns quickly reverse, falling by more than 50% by $t + 2$ (to 0.4%) and completely reversing by $t + 6$. The figure suggests that prices rise as investors flow into ETFs, but this price increase represents a temporary dislocation that will predictably reverse over the next 6 months. In other words, ETF investors are pushing prices above fundamental value, which causes them to underperform the very index they are trying to buy.

In many ways, ETFs are a natural laboratory for studying the joint impact of noise traders and arbitrage activity.[3] ETFs are similar to closed-end mutual funds in that they trade intra-day in the secondary market — any trader can buy a share of an ETF which is a claim to the ETF's underlying assets (e.g., the constituents of a stock index). However, ETFs differ from closed-end mutual funds because of the primary market mechanism in which pre-qualified parties, called authorized participants, may deliver the underlying assets and receive new shares when the price of the ETF price gets too high relative to its underlying. Similarly, authorized participants may deliver ETF shares and receive the underlying assets when the ETF price is too low relative to its underlying. Put simply, authorized participants create or remove shares in the ETF and trade

[3] We provide an extensive overview of the institutional details regarding the primary market for ETFs in Section 2.1.

them for the underlying assets to ensure the law of one price holds.

To understand the beauty of using share changes as a direct measure of arbitrage activity, consider the following. For any non-zero priced assets, exploiting an arbitrage opportunity implicitly requires a pairs trade, i.e., buy the cheap asset and sell the expensive asset. Thus, in other settings, measuring arbitrage requires (i) the ability to observe individual trades and (ii) the ability to match trades into arbitrage pairs. While not impossible, the task is difficult. However, in the world of ETFs, share creations and redemptions reflect the delivery of the cheaper asset in exchange for the more expensive asset.[4] Thus, ETFs' daily share changes are arguably the most direct measure of arbitrage activity available.

Moreover, ETFs are not simply a niche market; as such our findings have broad implications. First, with over two trillion dollars under management (BlackRock, 2014), ETFs are not a small asset class. Second, the vast majority of ETFs are not derivative products — they physically hold the underlying assets.[5] Third, ETFs are ingrained into nearly every asset market, both domestic and foreign.[6] Finally, ETFs are accessible to novice and professional investors alike.[7] Thus, our analysis of arbitrage activity in ETFs and their underlying assets touches nearly every corner of the global asset market.

Our data cover approximately 1,200 U.S. traded ETFs from January 2007 until December 2015. For each ETF we observe daily prices, share creation/redemption activities, net asset values, volumes, bid-ask spreads, underlying asset characteristics, and fund characteristics. In our tests, we focus on a sub-sample from January 2007 through December 2015 to ensure we have enough ob-

[4]Even in settings in which an authorized participant is hired by a client to create or redeem ETF shares, the authorized participant achieves the client's goals through whichever market, primary or secondary, is more favorable.

[5]ETFs typically replicate their predefined benchmarks by one of three methods: full replication, optimized replication, and derivative replication. Full replication is just that — an ETF physically holds all of its benchmarks' constituents in proportion to the benchmark weightings. Optimized replication is similar to full replication, but provides an ETF manager some discretion over which constituents to hold and which to avoid (perhaps due to limited liquidity). Derivative replication is achieved through futures or forward contracts on the fund's benchmark.

[6]With over 1,400 publicly traded ETFs in the United States, investors may construct portfolios with both domestic and international exposures and invest in everything from equities to real estate. For example, ETFs utilized nearly 100 unique Lipper objective codes in 2015. Lipper's objective codes are assigned based on the language that the fund uses in its prospectus to describe how it intends to invest. Lipper codes range from broad U.S domestic equities, e.g., "S&P 500 Index Objective Funds" to more exotic categories like "International Small-Cap Funds."

[7]ETFs are a popular investment choice within individual retirement plans, e.g., 401Ks, and also a popular investment for professional managers to "equitize" cash in their funds' benchmarks (Antoniewicz and Heinrichs (2014)).

servations when sorting funds into portfolios.[8] In most of our empirical exercises, we examine two cuts of the January 2007 through December 2015 data: (i) a sample in which funds' market capitalizations exceed $50MM, and (ii) a sub-sample of (i) in which funds are included after the first date on which one-half of the trading days within a month experienced some share creation/redemption activity.

Our empirical results examine the indirect effects of arbitrage activity. First, we consider arbitrage activity's ability to predict future asset returns. We sort ETFs into quintiles based on percent increases (decreases) in ETF shares outstanding at the weekly and monthly level. Portfolio returns are calculated and the differences between the extreme portfolios' abnormal returns are reported.[9] We include the results for both equal-weighted and value-weighted abnormal returns. The univariate sorts document statistically significant abnormal returns in the range of 5% to 13% per year. To examine the robustness of our results, we perform bivariate sorts and control for ETF fund-level short interest, ETF share bid-ask spreads, and ETF dollar volume. Across both weekly and monthly data, and equal and value-weighted abnormal returns, the return predictability from share changes appears strongest and most statistically significant in the ETFs which are the most liquid and easiest to trade.[10]

Our return predictability analysis suggests that when ETF fund size increases (i.e., when more shares are created), subsequent performance is poor and the opposite holds when ETF fund size decreases. In a sense, this means that ETF investors are poorly timing the market: they systematically buy when ETF prices are too high, relative to the value of the underlying assets, and they systematically sell when ETF prices are too low. To examine the return effect that this market mistiming has, we construct a *share-growth-adjusted* return that simultaneously accounts for changes in prices *and* share quantities. The measure is straightforward: we use each month's percent share

[8] While our analysis begins in 2007, the first exchange traded fund (SPY) in the United States was launched in January 1993 by State Street Global Advisors.

[9] We compute abnormal returns using Fama-French three-factor, Fama-French three-factor-plus-momentum, and five-factor models, however, we report only the abnormal returns using the Fama-French three-factor-plus-momentum model. The results for other models are similar.

[10] Our results are broadly consistent with the existing literature on transmission mechanisms which allow shocks to impact asset prices: short-horizon investors (who are also likely to be influenced by sentiment shocks) select into more liquid ETFs as in the models of Amihud and Mendelson (1986) and Constantinides (1986). As a result, liquid ETFs will be the ones most impacted by sentiment shocks.

change as a pseudo-leverage quantity for the subsequent month's return. For example, if an ETF's shares grow by three percent in June, we compute the July share-growth-adjusted return as

$$(1 + r_{\text{July}})(1.03) - 0.03.$$

For each fund, we compute a share-growth-adjusted return using a fund's series of one month total returns and its respective series of lagged share growth. To ascertain the significance of our share-growth-adjusted returns, we perform Monte Carlo simulations in which we perform the same calculation using a shuffled vector of the share growth. We find that share-growth-adjusted returns are systematically lower than what might occur by pure chance — on an equal weighted basis we find that more 9% of funds have have test statics smaller than -1.96 (by chance, one would expect 2.5%). On a market capitalization weighted basis, our results are even more pronounced with nearly 18% of the results falling below the -1.96 threshold. The results are also economically significant — the largest traded ETF, State Street Global Advisor's S&P500 ETF (i.e., SPY), has a share-growth-adjusted return of 4.96% (annualized) in our sample period as compared to the return on a single SPY share of 5.19%. Alternatively, one can compare the share-growth-adjusted return of 4.96% to the simulated, expected share-growth-adjusted return of 5.62%. The differences are between three and seven times larger than the fund's management fee of 9bps; in other words, ETFs may be significantly more costly than management fees suggest.

Our analysis adds to three bodies of research. First, we contribute to the growing literature on the relation between ETFs and other market outcomes. A number of papers argue that ETFs induce comovement between underlying assets (e.g., Baltussen, van Bekkum, and Da (2016), Da and Shive (2016)). Our analysis provides an additional insight; ETFs transmit temporary non-fundamental demand shocks to the assets, inducing time series return predictability. Recent empirical work shows that ETFs can serve as conduits that transmit non-fundamental market activity, i.e., investor sentiment, to their underlying assets. For example, Ben-David, Franzoni, and Moussawi (2014) document a relation between non-fundamental ETF price volatility and the prices of underlying assets.[11] Our work complements Ben-David et al. (2014) by focusing on the relation between

[11]Krause, Ehsani, and Lien (2013) also document a transmission of volatility from ETFs to the funds' underlying

arbitrage activity and return predictability. Notably, we show that conditioning on arbitrage activity leads to abnormal returns that cannot be explained by momentum or other canonical risk factors.

Second, we provide new insights into the dynamics of investor sentiment and arbitrage activity. Because most studies focus on observable mispricings, the existing literature has largely focused on the shortcomings of arbitrage activity (e.g., Pontiff (1996), Shleifer and Summers (1990), Shleifer and Vishny (1997), and Lamont and Thaler (2003)). We add new insights to this literature by focusing on observable arbitrage activity, rather than the absence of it. As a result, we document the effects that arbitrageur activity has on relative price efficiency and how it transmits to other assets.

Finally, our analysis also contributes to a recently revived discussion of Sharpe (1991)'s "arithmetic of active management" — the idea that active asset management as a whole *must* earn zero excess returns before fees. Sharpe's point is simple: passive market indexers earn exactly the market return (by definition), implying that any excess returns generated by active managers must come from *other* active managers. Thus, active managers, in total, must earn zero excess returns. While intuitive, the notion has come under recent scrutiny. Petajisto (2011) notes that passive index investing strategies will systematically under perform because they miss the positive performance associated with the *announcement* that a stock will be added to the index.[12] This underperformance by passive investors provides the opportunity for overperformance by active management, violating the arithmetic. More recently, Pedersen (2016) outlines additional violations of the arithmetic due to differential market access (e.g., passive investors' inability to participate in IPOs) and trading costs due to rebalancing. Our share-growth-adjusted return results suggest another significant way in which Sharpe's arithmetic may not hold. We show that ETFs tend to mechanically underperform their target index, i.e., an increase (decrease) in shares outstanding is subsequently accompanied with underperformance (overperformance). Thus, this underperformance provides an additional source of excess returns for active management. Overall, our results provide novel

assets.

[12]There is an extensive literature documenting positive abnormal returns after the announcement that a stock will be included in a major stock index, e.g., see Shleifer (1986).

evidence that trading by ETF investors generates predictable noise in returns, and this noise does not cancel out. In other words, noise traders can significantly impact asset prices.

The remainder of this paper proceeds as follows: Section 2 describes the basic market structure and briefly discusses the existing literature. Section 3 empirically tests the impact of arbitrage activity in ETFs on future returns using the cross-section. Section 4 examines the time-series effects of arbitrage activity. Section 5 concludes.

2 Background

The U.S. ETF market has grown dramatically over the last decade, total ETF assets have gone from $151 billion in 2003 to over $2 trillion in 2014 (BlackRock (2014)). Accordingly, academics, practitioners, and regulators, have all become increasingly interested in the structure of the ETF market and its impact on financial markets. In this section, we provide an overview of the institutional details regarding ETFs. We then provide a brief overview of the existing academic literature on ETFs and the impact of arbitrage activity.

2.1 Institutional Details

As discussed above, the ETF market has several unique features which make it an ideal laboratory for examining the impact of arbitrage activity. Perhaps most importantly, ETFs frequently trade on both a primary and secondary market. Investors can buy or sell ETFs on a secondary market just as they would buy or sell a stock. As a result, ETFs are generally viewed as highly liquid and transparent. However, ETFs are relatively unique in that they also regularly trade on a primary market.

Like a mutual fund, ETFs are a pooled investment vehicle which allows investors to buy a basket of assets at once.[13] However, unlike a mutual fund, shares in the ETF are added or removed on a primary market via the actions of third party traders called authorized participants (APs). APs, who are pre-qualified by the creator of an ETF, are allowed to exchange shares of the ETF

[13] Like mutual funds, most ETFs are formally registered with the SEC as investment companies under the Investment Company Act of 1940.

for shares of the underlying assets (and vice-versa). This process, which is designed to equilibrate supply and demand for shares in the ETF, allows APs to effectively enforce the law of one price in real time. For example, if an ETF price gets too high relative to the value of the underlying assets, an authorized participant would buy the underlying assets and trade them for newly created shares in the ETF. The AP would then sell these ETF shares on the secondary market and he would repeat the process until the law of one price held.

This market structure represents a unique opportunity to examine the impact of arbitrage activity. In particular, the market structure allows us to directly observe (i) trading by authorized participants, (ii) relative price efficiency between ETFs and the underlying assets, and (iii) the complete price process of both the ETF and the underlying asset at a daily frequency. As a result, we can use the ETF market to study the impact of arbitrage activity on price efficiency.

2.2 Existing Literature

Our paper is related to a growing literature on the impact of ETFs and index investing.[14] A number of papers argue that trading in ETFs can impact the properties of the underlying assets in the ETF portfolio. Specifically, several papers argue that ETFs can change the correlation structure of stock returns. Da and Shive (2016) show that ETF arbitrage can lead to comovement in equity returns. Similarly, Baltussen et al. (2016) show that serial correlation in equity returns goes from positive to negative after the introduction of ETFs around the world.[15]

Moreover, a number of papers argue that ETF trading can allow shocks to be transmitted to the underlying assets.[16] Israeli, Lee, and Sridharan (2015) present empirical evidence that stocks with greater ETF ownership experience relatively worse price efficiency.[17] In addition, Krause et al. (2013) argues that ETFs transmit volatility to stocks and Ben-David et al. (2014) shows that stocks

[14]In the interest of brevity, we omit a comprehensive literature review. For a helpful overview of the academic literature on ETFs, see Madhavan (2016) and Ben-David, Franzoni, and Moussawi (2016).

[15]Also related are the literatures examining the transmission effect mutual fund flows have on underlying asset prices (e.g., Coval and Stafford (2007), Lou (2012), Cella, Ellul, and Giannetti (2013), Hombert and Thesmar (2014), and Arif, Ben-Rephael, and Lee (2016)) and comovement induced by common institutional owners (e.g., Gao, Moulton, and Ng (2015)).

[16]More generally, the model (and empirical evidence) of Greenwood (2005) provides additional support for the demand-shock transmission hypothesis.

[17]In related work, Staer (2016) documents contemporaneous price pressure on the underlying stocks held by ETFs and subsequent reversals.

which are owned by ETFs experience increased volatility resulting from demand shocks. Consistent with this, Bhattacharya and O'Hara (2016) shows that ETFs can allow non-fundamental shocks to propagate into underlying asset prices and Malamud (2015) develops a general equilibrium model which shows that the creation/redemption mechanism may lead to higher volatility and momentum in the prices of underlying assets. However, the model also shows that the introduction of new ETFs may actually result in a reduction of volatility due to a demand substitution effect. Along these lines, Bessembinder (2015) explores the impact of predictable institutional order flow and its relation with asset prices. Bessembinder (2015) argues that predictable order flow from ETFs will have a relatively minor impact on asset prices, especially in the long-run since market participants will compete to provide liquidity and minimize price volatility arising from ETF re-balancing. Thus, ETFs will have a relatively minor impact on equity prices.[18]

In addition, there is a developed literature exploring *relative* mispricing between ETFs and their underlying assets. Engle and Sarkar (2006) shows that the premiums (discounts) between ETFs and their underlying assets are generally very small, and when they do exists, they last only a few minutes. In contrast, Fulkerson and Jordan (2012) examines ETF premiums or discounts and they find they can persist for five days. Moreover, Petajisto (2013) shows that ETF prices can be substantially different from their net asset values. While the average difference is only 14 basis points, the volatility of the difference is 66 basis points, which suggests there is substantial variation across funds.

There is also a nascent literature exploring *absolute* mispricing between ETFs and their underlying assets. Madhavan and Sobczyk (2014) develop a model of ETF price dynamics which shows how the creation/redemption mechanism in ETFs relates to deviations from an ETF's intrinsic value. In the model, price deviations arise either from price discovery or transitory liquidity effects. While the model is generally consistent with our findings, to the best of our knowledge, our paper is the first paper to empirically show that arbitrageur activity is linked to absolute mispricing.

More generally, our paper is related to the growing literature on the financial and real economic

[18]Etula, Rinne, Suominen, and Vaittinen (2016) show that regular flows into pensions plans at the end of each month induce some predictability into stock prices. However, their result is not about ETFs, per se, but rather the impact of predictable fund flows in general.

impacts of index investing. A number of papers show that index investing can lead to price effects (e.g., Shleifer (1986) and Kaul, Mehrotra, and Morck (2000)). More recently, Petajisto (2011) shows that the premium (discount) stocks receive for being added (subtracted) to an index can result in a hidden cost to investors who attempt to own the index, as they must buy stocks at a premium when they are added to the index and sell stocks at a discount when they are removed from the index. In addition, Brogaard, Ringgenberg, and Sovich (2016) show that index investing can impact the real economy by disrupting the signal in asset prices. The analysis shows that investments in commodity index ETFs can negatively impact firms which use commodities as an intermediate good. On the other hand, Appel, Gormley, and Keim (2016) show that passive index investing may actually lead to better outcomes for firms. The findings suggest that index funds exert a positive influence on firm performance in the long-term.

Finally, our analysis adds to a large literature examining violations of the law one price in financial markets.[19] Most similar to our work, several papers have examined deviations between closed-end funds' secondary market prices and their underlying net asset value.[20] Lee, Shleifer, and Thaler (1991), attributes mispricings to investor sentiment by showing that closed-end fund premia are correlated across funds and also linked to the holdings of individual investors. In a similar spirit, Pontiff (1996) attributes the observability of closed-end fund premia to limits of arbitrage, specifically showing that the premia are the largest in closed-end funds that are difficult for arbitrageurs to exploit. Our analysis builds upon Lee et al. (1991) and Pontiff (1996) by not only considering arbitrage opportunities but also the related activity by arbitrageurs. Thus, our strategy of using ETF share changes as a measure of arbitrage activity provides a new tool for future studies focused on violations of the law one price.

[19] In that sense, our work also relates to the large literature on the law of one price and the Fundamental Theorem of Finance (Dybvig & Ross, 1987). The Fundamental Theorem of Finance states that the absence of arbitrage implies the existence of a positive linear pricing rule that prices all assets. See also Cox and Ross (1975) and Harrison and Kreps (1979). Our work suggests that real world arbitrage opportunities exist, implying there may not be a positive linear pricing rule.

[20] For a survey on notable violations of the law of one price in financial markets, see Lamont and Thaler (2003).

3 Empirical Analyses

3.1 Data

To examine the theoretical predictions derived above, we combine data from Bloomberg, Compustat, CRSP, and Kenneth French's website. From Bloomberg, we get daily data on ETF share prices, ETF net asset values (NAV), ETF shares outstanding, ETF bid-ask spreads, and ETF trading volume.[21] Importantly, this data allows us to measure arbitrage activity via daily changes in share outstanding, which correspond to creations or redemptions in the ETF. Furthermore, Ben-David et al. (2014) suggest that Bloomberg provides the most accurate daily ETF data.

Each date, we calculate ETF premiums (discounts) as the difference between each ETF's price and its NAV. We then merge this data with information from CRSP including Lipper Codes and aggregate stock market returns. We also add monthly short interest data for each ETF from Compustat. Finally, to calculate a risk-adjusted measure of returns, we add information on the three factor (Fama & French, 1993), three factor plus momentum (Carhart, 1997), and five factors models (Hou, Xue, and Zhang (2016), Fama and French (2015)) from Kenneth French's website.

Table 1 displays a time-series count of the number of ETFs in our sample. As previously discussed, the ETF market has grown rapidly over the last decade, and by 2015, our sample includes more than 1,200 unique ETFs. To mitigate the impact of illiquidity and possible non-synchronous prices due to infrequent trading, we limit our sample to ETFs with at least $50 million in assets (this cutoff is consistent with a number of papers in the existing literature). The number of ETFs excluded using the $50 million threshold is approximately 25%, but only amounts to a reduction of $< 1\%$ of market cap. We also consider a sample of ETFs that are flagged as "mature" once they experience a month in which one-half of the trading days had some share creation/redemption activity. This is to ensure that we are analyzing ETFs in which both a primary and secondary market are vibrant. As shown in Table 1, this filter removes approximately half of the ETFs, but only reduces the total market cap by about 7%.[22]

[21] A number of ETFs have anomalous data on prices and shares outstanding that appear to be incorrect. Rather than winsorizing our data, we clean the data by removing the anomalies that are not verifiable via other data sources. Details of the cleaning method are available upon request.

[22] While this filter is not standard in the literature, we find similar results if we relax this filter so that we do not

Table 2 displays summary statistics for the sample of $50 million plus ETFs, as well as the mature sample of ETFs.[23] As expected, the mature sample is larger and generally experiences more trading and better liquidity; mature ETFs have more shares outstanding, more turnover, tighter bid-ask spreads, and more short interest. In Panel B of Table 2, we display information on the Lipper Categories of the ETFs. While the two samples are fairly similar, the mature ETFs tend to be more focused on equities and less focused on more exotic asset classes like bonds and international equities.

3.2 The Predictability Hypotheses

In this section, we consider whether or not arbitrage activity predicts asset returns. The null hypothesis is that arbitrage activity does not have predictive power. We start by creating portfolios each period based on last period's $ShareChange$ in percent. In other words, our sorts are designed to see whether past ETF arbitrage activity by authorized participants is related to future ETF performance, controlling for risk factors. Under the null hypothesis, we do not expect percent changes in ETF shares outstanding to be associated with ETF performance. In all of our portfolio sorts, we sort based on characteristics at a weekly/monthly level, preventing time trends and differences in sample size from driving our results.[24,25] While our analysis is qualitatively unchanged when using raw returns, three factor abnormal returns, or five factor abnormal returns, we report only the results using a four-factor model.

The results of the univariate portfolio sort are shown Table 3 and we report both equal-weighted and value-weighted results. Portfolio 5 minus portfolio 1 yields statistically significant abnormal returns in Panel A for both subsamples when using equal-weighted returns. On an annualized basis, these returns range from 5.8% to 7.3%. The results are not statistically significant at the

exclude such a large number of ETFs.

[23] The entire sample of ETFs is omitted for the sake of brevity as the $50 million plus sample is representative of the entire sample

[24] We do not perform analysis at a daily frequency intentionally. Share creation/redemption activity accounting varies across ETFs with some funds using $T+1$ accounting (meaning they register the share creation activity the day after it occurs) while others using T accounting. Moreover, within funds the accounting has changed. A fund's change from $T+1$ to T accounting is not public. See Staer (2016) for additional details.

[25] In unreported analysis, we perform sorts at a quarterly frequency. The quarterly results are not statistically significant, suggesting that the mispricings we observe are corrected over horizons shorter than one quarter.

weekly level when using value-weighted returns, but the coefficients share the same sign. It is worthwhile to mention a bias introduced with value-weighted returns — the largest ETFs typically have the same underlying assets. For example, in 2015 three of the top six largest ETFs tracked the S&P500. Thus, any analysis done using value-weighted returns is going tilt the analysis towards a small set of funds that follow generally the same assets. Given the potential short-comings of value-weighted and equal-weighted returns, we report the results for both value-weighted and equal-weighted hereafter.

In Panel B of Table 3, the results are reported at the monthly level. Results are statistically significant for both portfolios using equal-weighted returns. The economic magnitude in the $50 million plus sample is about 52bps (significant at 1% level) per month which annualizes to about 6.4%. In the mature ETF sample, the monthly abnormal return is significantly larger at about 101bps (significant at 1% level) per month which annualizes to 12.8%. In the analysis with value-weighted returns, the two samples give abnormal returns of about 43bps (significant at 10% level) and 51 bps (significant at 5% level). Taken together, the coefficients across weighting schemes and samples are stable and demonstrate economically meaningful abnormal returns.

3.3 Robustness

It possible that our results up until this point are merely documenting characteristics of ETFs and future ETF performance. For example, it is possible that the abnormal returns observed in the univariate sorts are symptomatic of wide bid-ask spreads or that the results are being driven by small, illiquid ETFs. Accordingly, to control for these possible confounding influences, we examine dual sorts, using the sorting variable $ShareChange$ after first using one of the following $\{ShortIntPct, BidAskSpread, Volume\}$, where $ShortIntPct$ is calculated as the percent of shares held short in each ETF, each month, from Compustat expressed as a percentage of shares outstanding from CRSP, $BidAskSpread$ is calculated as an EFT's average ask minus bid divided by the midpoint, and $Volume$ is the average dollar volume of an ETF.

The first set of bivariate results sort on $ShareChange$ and $ShortIntPct$. Recall, larger values of $ShortIntPct$ are typically associated with better ETF liquidity. The results using equal-weighted returns are reported in Table 4. The sort is performed on the mature sample and all abnormal returns are computed from a four factor model. In Panel A, weekly data sorts are reported. The differences between the fifth and first portfolios are largest in the three highest quintiles of $ShortIntPct$, with the differences in the third and fourth quintiles being significant. In Panel B, monthly results are reported. The difference in portfolio 5 minus portfolio 1 returns is largest in the three largest portfolios of $ShortIntPct$ and all are statistically significant at the 1% level. The results suggest that the return predictability coming from arbitrage activity is strongest in the more easy-to-short ETFs (i.e., ETFs which have relatively lower limits to arbitrage). In Table 5, a similar analysis is performed sorting on $ShortIntPct$, however, returns are value-weighted. Nevertheless, the results are consistent, yielding the largest differences in portfolio returns in the ETFs with the largest values of $ShortIntPct$.

The next set of bivariate results sort on $ShareChange$ and $BidAskSpread$. The results for equal-weighted returns are reported in Table 6 with Panel A reporting weekly portfolio abnormal returns and Panel B reporting monthly abnormal returns. At the weekly level, the strongest and most statistically significant difference in 5-1 portfolios along the $ShareChange$ dimension occur in the portfolios with tight bid-ask spreads. To put an economic meaning on the coefficients, exploiting the 5-1 abnormal returns of -0.30% and -0.39% in the two smallest bid-ask spread portfolios would amount to annualized returns of approximately 17% and 23% annually. The monthly results are generally in line with the weekly results, albeit slightly weaker economically. The results for value-weighted returns are reported in Table 7 and are consistent with the equal-weighted results Table 6, although the economic magnitudes in both the weekly and monthly frequencies are slightly smaller.

Last, we perform bivariate sorts on $ShareChange$ and $Volume$. The results for equal-weighted returns appear in Table 8 with Panel A reporting weekly returns and Panel B reporting monthly returns. Consistent with the other bivariate sorts, the difference between portfolio 5 and portfolio 1 on the $ShareChange$ dimension is strong in the highest volume ETFs, which are presumably more liquid. The value-weighted returns reported in Table 9 yield similar monthly return results as equal-weighted returns. Interestingly, the lowest volume ETFs also show large differences in fifth minus first portfolio returns at the monthly level. This suggests multiple mechanisms may be leading to arbitrage-activity-based return predictability.

Taken together, our robustness checks, while noisy, indicate return predictability from ETF arbitrage activity, even when controlling for other factors. Importantly, the results appear to be the strongest in the most liquid ETFs. Specifically, in most specifications, the strongest results (both economically and statistically) are in the most liquid subset of ETFs. Furthermore, most sorts, even when statistically insignificant, exhibit the correct sign on the difference, i.e., portfolios with higher share changes tend to be associated with lower returns.

4 Share-Growth-Adjusted Return Analysis

Next, we examine the relation between share creation and future returns on a fund-by-fund basis. To do so, we consider monthly returns from each ETF and lagged monthly share creations. Specifically, for each ETF, we take its return series $\{r_1, \ldots, r_T\}$ and its one-period-lagged share growth series $\{g_0, \ldots, g_{T-1}\}$ and perform the following calculation,

$$R \equiv \prod_{\tau=1}^{T} \big((1 + r_\tau)(1 + g_{\tau-1}) - g_{\tau-1}\big). \tag{1}$$

Thus, R is a pseudo portfolio return over the T periods in the sample period — the analytic expression of R captures the notion that share creations and redemptions have a leverage-like effect on the total return of the funds total portfolio. For example, if all ETF shares were collectively held by a representative investor, share creations would provide the investor with greater exposure to the ETF's benchmark and redemptions would reduce exposure. We convert R into an annual return according to:

$$(1 + \overline{r}) = R^{12/T}, \tag{2}$$

and we refer to $\overline{r}$ as the *share-growth-adjusted return.*

If share creations and redemptions are uncorrelated with future returns, then $\overline{r}$ is purely random. To examine whether $\overline{r}$ is random or not, we perform a Monte Carlo simulation of 1,000,000 paths for each ETF. In a given Monte Carlo path i, the vector $\{g_0, \ldots, g_{T-1}\}$ is shuffled into a new vector

$\{\hat{g}_{0,i}, \ldots, \hat{g}_{T-1,i}\}$. For example, $\hat{g}_{0,i}$ in the shuffled vector may correspond to g_{T-1} in the actual share growth series. Subsequently within each path i, the calculation,

$$R_i^{MC} \equiv \prod_{\tau=t}^{T} \big((1+r_\tau)(1+\hat{g}_{\tau-1,i}) - \hat{g}_{\tau-1,i}\big), \tag{3}$$

is performed and its corresponding annual return $\overline{r}_i^{MC}$ is calculated,

$$(1+\overline{r}_i^{MC}) = (R_i^{MC})^{12/T}. \tag{4}$$

The Monte Carlo paths yield a distribution which is subsequently used to test the statistical significance of the *share-growth-adjusted return* $\overline{r}$.[26]

Using our mature ETF sample we examine the statistical significance of $\overline{r}$ for each fund using monthly ETF returns and monthly share changes. We restrict our analysis to ETFs for which we have at least 36 months of data. Table 10 provides a breakdown of the statistical significance of $\overline{r}$ in the mature ETF sample. In Panel A, the percentages of funds with $\overline{r}$ that are statistically different than the expected fee implied by the Monte Carlo simulations are bucketed into those with t-stats less than -1.96 and those with t-stats greater than 1.96.[27] As a reference, we also provide a breakdown of the t-stats, while statistically insignificant, below zero and above zero. The percentage of funds with t-stats below -1.96 is 9.86%. One would expect to see approximately 2.5% of the observations in that tail by chance. Thus, with over 400 ETFs in our mature sample, we find just under four times as many funds in the left-tail than would be expected if share-growth-adjusted returns are purely random. Conversely, 5.63% of the observations are significant in the right-tail, which is a little more than two times what one would expect if share-growth-adjusted returns are purely random. The test suggests a non-zero covariance between share growth and subsequent returns, which may be positive or negative, with there being almost twice as many negative (and significant) share-growth-adjusted returns as positive.

[26] Due to some periods of large share creation changes and some periods of large return swings, some Monte Carlo simulation paths result in $R_i^{MC} < 0$. In these settings, we set $R_i^{MC} = 0$ as an absorbing state, i.e., the ETF goes out of business.

[27] With 1,000,000 Monte Carlo paths, a test statistic of 1.96 corresponds to a two-tailed p-value of 0.05.

The results are depicted visually in Figure 2. The horizontal axis in the figure represents the difference between $\overline{r}$ and its expected value based on the simulated distribution. The tick marks are in basis points and, for the sake of presentation, truncated at -500 bps and 500 bps. The vertical axis represents the p-value of each observation and the axis runs from 0 to 0.05. Thus, only observations of $\overline{r}$ that are statistically different from $E[\overline{r}]$ are plotted, with the strongest statistical significance appearing closest to the horizontal axis. The different size circles are related to each fund's market capitalization, with our measure (detailed shortly) of the largest ETFs depicted with the largest circles. Visually, it appears that the majority of the statistically significant observations of $\overline{r}$ are negative, suggesting a negative covariance between lagged share creation and future returns.

In the previous paragraph, we eluded to a measure of market capitalization. The ETF market has been characterized by both rapid growth and the rapid introduction of new funds. As such, weighting by market capitalization can be problematic. For example, if one were to use average market capitalization over the sample period, the rapid growth in assets under management would bias the weighting scheme towards new, larger ETFs. As such, in an effort to minimize time effects, we create a measure of market capitalization termed *average market capitalization share*. For each fund j, the measure is computed as,

$$\overline{MC}_j = \frac{\sum_{\tau}^{T} \left(\frac{MC_{j,\tau}}{\sum_{k}^{J_\tau} MC_{k,\tau}} \right)}{T}. \tag{5}$$

The term within parenthesis in (5) represents the fraction of the ETF market that fund j has in year τ of the sample. Due to the introduction of new ETFs after 2007, the number of funds in each year τ may change in our sample from 2007 to 2016. As such, the summation takes into consideration only the funds that existed in year τ. A series of market capitalization shares is calculated for fund j and then the average is computed for the T years that fund is in the sample. The end result, $\overline{MC}_j$, represents the average share of the ETF market held by the fund.

Returning to Table 10, we perform an analysis in Panel B that computes a market capitalization weighted percentage of observations in each tail. The weights are determined by each fund's average market capitalization share outlined in (5). On a weighted basis, the percentage of results that end up in the left-tale, i.e., those with test statistics smaller than -1.96, is approximately 19.29% of the sample. The number is twice that of when observations are equal-weighted in Panel A. Conversely, the percentage of funds in the right-tale is just under 5.00%, which is smaller than the equal-weighted counterpart in Panel A. The results suggest a skew of share-growth-adjusted returns that are largely negative, statistically significant, and exacerbated by fund size. Moreover, while statistically insignificant, 42.64% of the observations are below zero (and above -1.96) while 33.33% are above zero (and below 1.96), suggesting that the results are not symmetric around zero when accounting for fund size.

To examine the robustness of our results, we perform three additional tests. In Table 11, we restrict our sample to only those ETFs that existed before the start date in 2007. Panel A presents equal-weighted buckets based on test statistics. Again, both the right and left tails exhibit a higher frequency of funds than would be expected by chance, with each tail possessing 8.11% of the funds. Panel B displays the analysis with value-weighted buckets. The value-weighted percentage of funds below a test statistic of -1.96 is 23.41%, nearly one quarter of the sample. Conversely, only about 3% of the value-weighted observations exceed a test statistic of 1.96. In Panel B, the statistically insignificant results again exhibit a negative skew around zero with 46.04% of the sample falling in the (-1.96,0) region.

In Table 12, we provide the complementary analysis. We restrict our sample to only those ETFs that were introduced after the start date in 2007. Panel A again represents the equal-weighted analysis and Panel B represents the value-weighted analysis. Both panels exhibit the same trends as were discussed from Tables 10 and 11; there exists a significant fraction of funds with negative and statistically significant test statistics in both the equal-weighted and value-weighted samples. Finally, in Table 13 we restrict the sample to just the 100 largest ETFs by average market capitalization share. Broadly speaking, this subset of ETFs is characterized by tight bid-ask spreads and large average dollar volume. The results mirror those outlined earlier: a significant fraction of the ETFs appear in the left-hand tail on an equal-weighted basis and an even larger fraction appears on a value-weighted basis.

The share-growth-adjusted returns studied in this section are consistent with the results in Section 3.2: in a given ETF, share creations and redemptions covary with the fund's future performance. Furthermore, the covariance between share creations/redemptions and subsequent performance is largely negative and exacerbated in the largest ETFs. While the analysis opens discussions on several important issues, there is one specific discussion we believe is worth highlighting. Our share-growth-adjusted return accounts for the dynamic performance of an ETF given that both prices and quantities (i.e., assets under management) fluctuate. In comparison, a typical return analysis is performed by examining the return on a given share of an ETF that is never redeemed. Focusing on a single share's return does not account for the ETF exposure (i.e., size) changing. If changes in quantities and changes in prices are unrelated, then examining the return on a single share does not appear problematic. However, our analysis suggests that the changes in quantities and prices are related. Therefore, studies examining ETF returns may need to control for both changes in prices (as typical stock return studies do) *and* changes in quantities.

For example, while we are agnostic towards the social welfare consequences of our results, accounting for both changes in quantities and prices has implications for ETF investors. If all ETF shares were held by a single representative investor, her performance would be closer to that of our share-growth-adjusted return than the performance on a single ETF share. Obviously ETFs are not held by one individual, but allocating returns is a zero sum game. Thus, while a buy-and-hold investor may earn a return consistent with that on a single share, other higher frequency traders must absorb the residual performance. To put things in perspective, consider the ETF SPY. SPY is the State Street Global Advisor's ETF that mimics the S&P500 index via full replication and also the largest ETF in our sample. SPY has a share-growth-adjusted return $\overline{R}$ of 4.96% which differs from its simulated expected value $E[\overline{R}]$ of 5.62% by 66 bps basis points.[28] Thus, while SPY's management fee is 9 bps per anum, our analysis suggests that investors bear an additional indirect cost that is nearly seven times larger.

Another implication of our analysis applies broadly to asset pricing tests. When conducting

[28] As an additional point of reference, the S&P500's annualized total return over the sample horizon (which does not account for share creations and redemptions) was 5.19% which is 23 basis points higher than SPY's share-growth-adjusted return.

cross sectional asset pricing test using a typical four factor model, it might be important to consider an ETF's return loading on the given factor *and* share change loading. That is, because ETF share quantities and ETF share prices covary together, an understanding of how asset pricing factors jointly affect quantities and prices may be of interest. This is a topic of future research.

5 Conclusion

Theoretical asset pricing models often assume the existence of noise traders, yet measuring them poses an empirical challenge. As a result, there are relatively few empirical papers examining how noise traders and arbitrageurs impact the formation of asset prices. In this paper, we exploit a novel data set to examine arbitrage activity by authorized participants in the market for ETFs. The data provides a unique measure of noise trading and subsequent arbitrageur activity. Our results show that when a large amount of money flows into ETFs it leads to distortion in the ETF's price. Thus, ETF investors tend to systematically buy (and sell) assets at the wrong price. In other words, ETF investors are a real-world example of noise traders. We then show that arbitrage activity does indeed lead to a price correction. Following a flow shock, authorized participants create (or redeem) shares in the ETF, leading to subsequent price reversals.

Our results make several contributions. First, we show that noise trading does exist and, importantly, noise traders do not always cancel each other out, they impact asset prices. In the process, we contribute to the growing literature which examines the impact of ETFs. Second, we provide novel evidence on the impact of arbitrage activity. While many papers focus on the absence of arbitrage, we provide novel evidence on observable arbitrage activity. Finally, we provide new evidence that active managers may be able to earn excess returns, as a group, because passive indexers systematically underperform. As such, our results provide a new counter-point to Sharpe's well known arithmetic about active management. Overall, our results show that noise traders and arbitrageurs exert a powerful impact on the dynamics of asset prices.

References

Amihud, Y., & Mendelson, H. (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics*, *17*(2), 223–249.

Antoniewicz, R. S., & Heinrichs, J. (2014). Understanding exchange-traded funds: How etfs work. *Jane, Understanding Exchange-Traded Funds: How ETFs Work (September 30, 2014)*.

Appel, I., Gormley, T. A., & Keim, D. B. (2016). Passive investors, not passive owners. *Journal of Financial Economics*, *forthcoming*.

Arif, S., Ben-Rephael, A., & Lee, C. (2016). What drives short-sale strategies? new evidence from mutual fund trades. *New Evidence from Mutual Fund Trades (August 5, 2016)*.

Baltussen, G., van Bekkum, S., & Da, Z. (2016). Indexing and Stock Market Serial Dependence Around the World. *Working Paper*.

Ben-David, I., Franzoni, F., & Moussawi, R. (2014). Do ETFs Increase Volatility? *Working Paper*.

Ben-David, I., Franzoni, F., & Moussawi, R. (2016). *Exchange traded funds (etfs)* (Tech. Rep.). National Bureau of Economic Research.

Bessembinder, H. (2015). Predictable ETF Order Flows and Market Quality. *Working Paper*.

Bhattacharya, A., & O'Hara, M. (2016). Can ETFs Increase Market Fragility? Effect of Information Linkages in ETF Markets. *Working Paper*.

BlackRock. (2014). Etp landscape, industry highlights. *BlackRock Report*.

Brogaard, J., Ringgenberg, M., & Sovich, D. (2016). The Economic Impact of Index Investing. *Working Paper*.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, *52*(1), 57-82.

Cella, C., Ellul, A., & Giannetti, M. (2013). Investors' horizons and the amplification of market shocks. *Review of Financial Studies*, hht023.

Constantinides, G. M. (1986). Capital market equilibrium with transaction costs. *The Journal of Political Economy*, 842–862.

Coval, J., & Stafford, E. (2007). Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics*, *86*(2), 479–512.

Cox, J. C., & Ross, S. A. (1975). A Survey of Some New Results in Financial Option Pricing Theory. *Journal of Finance*, *31*, 383-402.

Da, Z., & Shive, S. (2016). Exchange-traded funds and equity return correlations. *Working Paper*.

Dybvig, P. H., & Ross, S. A. (1987). arbitrage. In J. Eatwell, M. Milgate, & P. Neuman (Eds.), *The new palgrave: a dictionary of economics* (p. 100106). London: Macmillan.

Engle, R., & Sarkar, D. (2006). Premiums-Discounts and Exchange Traded Funds. *The Journal of Derivatives*, *13*(4), 27–45.

Etula, E., Rinne, K., Suominen, M., & Vaittinen, L. (2016). Dash for Cash: Month-End Liquidity Needs and the Predictability of Stock Returns. *Working Paper*.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*, 3-56.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1-22.

Fulkerson, J. A., & Jordan, B. D. (2012). Reading Tomorrow's Newspaper: Predictability in ETF Returns. *Working Paper*.

Gao, G., Moulton, P. C., & Ng, D. T. (2015). Institutional ownership and return predictability across economically unrelated stocks. *Johnson School Research Paper Series*(15-2013).

Greenwood, R. (2005). Short-and long-term demand curves for stocks: theory and evidence on the dynamics of arbitrage. *Journal of Financial Economics*, *75*(3), 607–649.

Gromb, D., & Vayanos, D. (2010). Limits of arbitrage. *Annual Review of Financial Economics*, *2*(1), 251–275.

Harrison, J. M., & Kreps, D. M. (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, *20*, 381-408.

Hombert, J., & Thesmar, D. (2014). Overcoming limits of arbitrage: Theory and evidence. *Journal of Financial Economics*, *111*(1), 26–44.

Hou, K., Xue, C., & Zhang, L. (2016). A comparison of new factor models. *Working Paper*.

Israeli, D., Lee, C., & Sridharan, S. A. (2015). Is there a dark side to exchange traded funds (etfs)? an information perspective. *An Information Perspective (July 1, 2015)*.

Kaul, A., Mehrotra, V., & Morck, R. (2000). Demand curves for stocks do slope down: New evidence from an index weights adjustment. *Journal of Finance*, *55*, 893-912.

Keynes, J. M. (1936). *The general theory of employment interest and money* (Tech. Rep.).

Krause, T., Ehsani, S., & Lien, D. (2013). Exchange traded funds, liquidity, and market volatility.

Working Paper.

Lamont, O. A., & Thaler, R. H. (2003). Anomalies: The law of one price in financial markets. *The Journal of Economic Perspectives*, *17*(4), 191–202.

Lee, C., Shleifer, A., & Thaler, R. H. (1991). Investor sentiment and the closed-end fund puzzle. *The Journal of Finance*, *46*(1), 75–109.

Lou, D. (2012). A flow-based explanation for return predictability. *Review of Financial Studies*, *25*(12), 3457–3489.

Madhavan, A. (2016). Exchange-Traded Funds: An Overview of Institutions, Trading, and Impacts. *Annual Review of Financial Economics*, *6*, 311-341.

Madhavan, A., & Sobczyk, A. (2014). Price Dynamics and Liquidity of Exchange-Traded Funds. *Working Paper*.

Malamud, S. (2015). A Dynamic Equilibrium Model of ETFs. *Working Paper*.

Pedersen, L. H. (2016). Sharpening the arithmetic of active management. *Browser Download This Paper*.

Petajisto, A. (2011). The index premium and its hidden cost for index funds. *Journal of Empirical Finance*, *18*, 271-288.

Petajisto, A. (2013). Inefficiencies in the Pricing of Exchange-Traded Funds. *Working Paper*.

Pontiff, J. (1996). Costly arbitrage: Evidence from closed-end funds. *The Quarterly Journal of Economics*, 1135–1151.

Sharpe, W. F. (1991). The arithmetic of active management. *Financial Analysts Journal*, *47*(1), 7–9.

Shleifer, A. (1986). Do demand curves for stocks slope down? *Journal of Finance*, *41*, 579-590.

Shleifer, A., & Summers, L. H. (1990). The noise trader approach to finance. *The Journal of Economic Perspectives*, *4*(2), 19–33.

Shleifer, A., & Vishny, R. W. (1997). The limits of arbitrage. *The Journal of Finance*, *52*(1), 35–55.

Staer, A. (2016). Fund flows and underlying returns: The case of etfs. *Available at SSRN 2158468*.

Figure 1: Cumulative Abnormal ETF Returns Around Top-Quintile Share Creation Events. ETF returns are adjusted using monthly, value-weighted average ETF returns. Quantiles are calculated over the entire sample period, from 2007 through 2015, and the sample includes all mature ETFs.

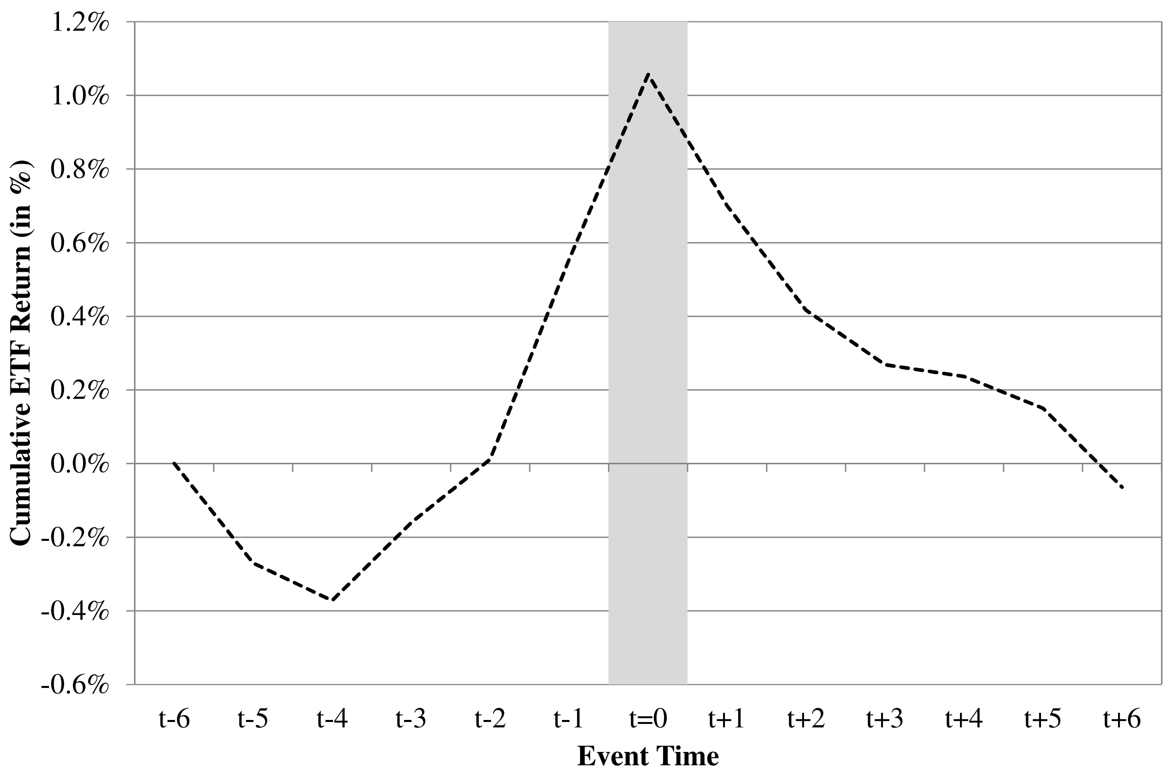

Figure 2: Scatter plot of all mature ETFs weighted by average market capitalization share: the figure depicts the difference between realized returns $\bar{r}$ and $\mathrm{E}[\bar{r}]$ based on Monte Carlo simulation. The horizontal axis represents $\bar{r} - \mathrm{E}[\bar{r}]$ in basis points. The vertical axis is the two-tailed p-value. The sizes of the scatter plot circles are determined by funds average market capitalization share.

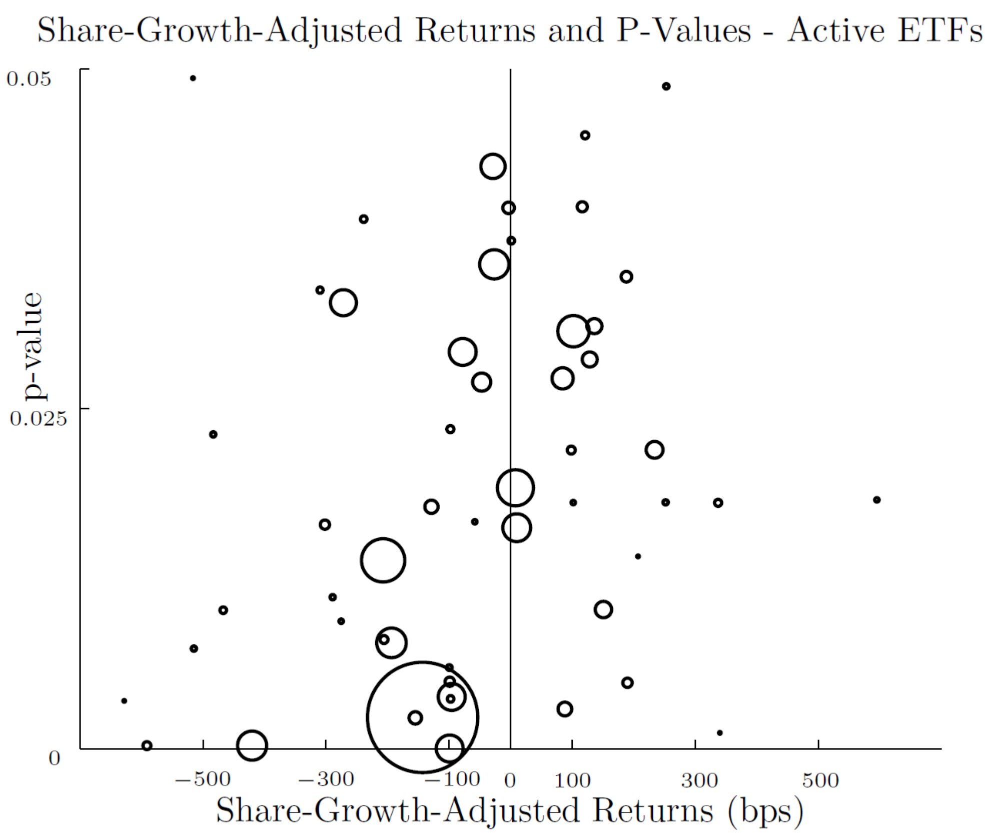

Table 1: Number of ETFs per year. ETFs are considered mature from the first month in which creation or redemption activity was reported on at least 50% of trading days.

	All ETFs		$50M+ ETFs		Mature ETFs	
Year	Number	Market Cap	Number	Market Cap	Number	Market Cap
2007	277	$396	223	$394	82	$328
2008	423	$546	334	$545	122	$466
2009	517	$484	413	$483	180	$434
2010	601	$661	469	$658	207	$581
2011	759	$964	601	$961	270	$872
2012	885	$1,100	680	$1,097	314	$1,011
2013	987	$1,355	754	$1,352	347	$1,243
2014	1,091	$1,554	860	$1,550	383	$1,412
2015	1,199	$1,774	956	$1,771	418	$1,595

Table 2: ETF Summary Characteristics. ETFs are considered mature from the first month in which creation or redemption activity was reported on at least 50% of trading days.

	\$50M+ ETFs	Mature ETFs
Average ETF Characteristics		
Shares Outstanding (millions)	31.3	62.3
Average Monthly Volume (millions)	36	79
Average Monthly Volume (percentage of shares out)	97.6%	167%
ETF Market Capitalization (billions)	\$1.7	\$3.4
Bid-Ask Spread	0.19%	0.09%
Short Interest Percentage	6.3%	10.9%
Lipper Category Percentages		
Bonds	11.9%	9.7%
Commodities	2.3%	2.7%
General Equities	41.9%	43.6%
Mixed Assets	0.5%	0.1%
International	20.5%	19.0%
Municipal Bonds	2.0%	1.5%
Sector-Based Equities	20.9%	23.5%

Table 3: Portfolio alphas based on ETF arbitrage activity. ETFs are sorted, per period (week or month), based on the prior period's aggregate creation / redemption activity into one of five portfolios. Portfolio returns are calculated based on all ETF-period observations, and these portfolio returns are regressed on the market, small-minus-big, high-minus-low and momentum factors. The last three columns present results from testing the difference in returns between the portfolios with the highest creation and highest redemption activity. ETFs are considered mature from the first month in which creation or redemption activity was reported on at least 50% of trading days.

	Portfolios Sorted on *ShareChange*							
	1(low)	2	3	4	5(high)	$5-1$	std err	p-value
Panel A: Weekly Portfolio Alphas (in percent)								
$50M+ ETFs								
Equal-Weighted	-0.004	-0.029	-0.004	-0.050	-0.112	-0.108	0.051	0.035
Value-Weighted	0.072	0.026	0.003	-0.001	0.027	-0.045	0.051	0.380
Mature ETFs								
Equal-Weighted	-0.021	-0.056	-0.019	-0.026	-0.156	-0.136	0.075	0.070
Value-Weighted	0.087	0.029	-0.011	0.024	-0.005	-0.092	0.059	0.122
Panel B: Monthly Portfolio Alphas (in percent)								
$50M+ ETFs								
Equal-Weighted	-0.075	-0.239	-0.140	-0.136	-0.596	-0.521	0.216	0.017
Value-Weighted	0.237	-0.119	0.182	0.013	-0.187	-0.425	0.238	0.076
Mature ETFs								
Equal-Weighted	0.006	-0.271	-0.165	-0.219	-1.006	-1.012	0.283	0.000
Value-Weighted	0.137	-0.011	0.086	0.018	-0.369	-0.506	0.239	0.035

Table 4: Equal-weighted portfolio alphas based on ETF arbitrage activity and ETF short interest. ETFs are first sorted, per period (week or month), into five portfolios based on the prior period's ETF short interest. ETFs are then sorted, within each short-interest portfolio, based on the prior period's aggregate creation / redemption activity, resulting in twenty five total portfolios. Portfolio returns are calculated based on all ETF-period observations, and these portfolio returns are regressed on the market, small-minus-big, high-minus-low and momentum factors. The last three columns and last three rows of each panel present results from testing the difference in returns between the highest and lowest portfolios.

Panel A: Weekly Portfolio Alphas (in percent): Mature ETFs

	Portfolios Sorted on *ShareChange*							
Portfolios Sorted on *ShortIntPct*	1(low)	2	3	4	5(high)	$5-1$	std err	p-value
1(low)	-0.398	-0.252	-0.334	-0.225	-0.158	0.240	0.175	0.175
2	-0.168	-0.289	-0.203	-0.077	-0.123	0.045	0.209	0.830
3	-0.116	-0.382	0.017	-0.332	-0.763	-0.648	0.344	0.063
4	0.457	-0.196	-0.272	-0.540	-0.715	-1.172	0.402	0.004
5(high)	0.374	-0.524	-0.127	0.207	-0.284	-0.657	0.428	0.128
$5-1$	0.772	-0.272	0.208	0.432	-0.125			
std err	0.315	0.246	0.230	0.261	0.338			
p-value	0.016	0.271	0.369	0.102	0.712			

Panel B: Monthly Portfolio Alphas (in percent): Mature ETFs

	Portfolios Sorted on *ShareChange*							
Portfolios Sorted on *ShortIntPct*	1(low)	2	3	4	5(high)	$5-1$	std err	p-value
1(low)	-0.426	-0.214	-0.218	0.131	-0.067	0.359	0.208	0.086
2	-0.139	-0.326	-0.308	-0.020	-0.426	-0.287	0.372	0.441
3	0.251	-0.587	0.023	-0.043	-1.044	-1.295	0.499	0.010
4	0.101	0.041	-0.306	-0.614	-1.831	-1.932	0.639	0.003
5(high)	0.467	-0.218	-0.410	0.064	-1.162	-1.629	0.573	0.005
$5-1$	0.893	-0.005	-0.192	-0.067	-1.095			
std err	0.501	0.258	0.246	0.241	0.347			
p-value	0.076	0.986	0.435	0.782	0.002			

Table 5: Value-weighted portfolio alphas based on ETF arbitrage activity and ETF short interest. ETFs are first sorted, per period (week or month), into five portfolios based on the prior period's ETF short interest. ETFs are then sorted, within each short-interest portfolio, based on the prior period's aggregate creation / redemption activity, resulting in twenty five total portfolios. Portfolio returns are calculated based on all ETF-period observations, and these portfolio returns are regressed on the market, small-minus-big, high-minus-low and momentum factors. The last three columns and last three rows of each panel present results from testing the difference in returns between the highest and lowest portfolios.

Panel A: Weekly Portfolio Alphas (in percent): Mature ETFs

	Portfolios Sorted on *ShareChange*							
Portfolios Sorted on *ShortIntPct*	1(low)	2	3	4	5(high)	$5-1$	std err	p-value
1(low)	-0.550	-0.164	-0.127	0.009	-0.210	0.340	0.194	0.083
2	-0.092	-0.368	-0.247	-0.053	-0.007	0.085	0.215	0.694
3	0.068	-0.505	0.264	0.036	-0.561	-0.629	0.361	0.085
4	0.739	-0.099	-0.183	-0.217	0.145	-0.593	0.480	0.220
5(high)	-0.148	-0.518	0.049	0.010	0.054	0.202	0.274	0.463
$5-1$	0.402	-0.354	0.176	0.001	0.264			
std err	0.229	0.239	0.207	0.195	0.246			
p-value	0.082	0.143	0.396	0.997	0.286			

Panel B: Monthly Portfolio Alphas (in percent): Mature ETFs

	Portfolios Sorted on *ShareChange*							
Portfolios Sorted on *ShortIntPct*	1(low)	2	3	4	5(high)	$5-1$	std err	p-value
1(low)	-0.267	0.007	0.067	0.106	0.062	0.329	0.266	0.218
2	0.218	0.003	-0.201	-0.085	-0.347	-0.565	0.382	0.141
3	0.152	-0.084	0.307	0.157	-0.409	-0.562	0.563	0.319
4	0.078	-0.018	-0.169	-0.180	-0.731	-0.809	0.693	0.244
5(high)	0.501	0.160	0.031	0.154	-0.722	-1.224	0.379	0.001
$5-1$	0.768	0.153	-0.036	0.048	-0.784			
std err	0.369	0.283	0.229	0.221	0.280			
p-value	0.039	0.589	0.875	0.828	0.006			

Table 6: Equal-weighted portfolio alphas based on ETF arbitrage activity and bid-ask spread. ETFs are first sorted, per period (week or month), into five portfolios based on the prior period's bid-ask spread. ETFs are then sorted, within each bid-ask-spread portfolio, based on the prior period's aggregate creation / redemption activity, resulting in twenty five total portfolios. Portfolio returns are calculated based on all ETF-period observations, and these portfolio returns are regressed on the market, small-minus-big, high-minus-low and momentum factors. The last three columns and last three rows of each panel present results from testing the difference in returns between the highest and lowest portfolios.

Panel A: Weekly Portfolio Alphas (in percent): Mature ETFs

	Portfolios Sorted on *ShareChange*							
Portfolios Sorted on *BidAskSpread*	1(low)	2	3	4	5(high)	$5-1$	std err	p-value
1(low)	0.179	-0.017	-0.008	-0.015	-0.122	-0.302	0.113	0.008
2	0.161	-0.067	-0.017	-0.078	-0.233	-0.394	0.131	0.003
3	-0.229	-0.028	-0.025	0.002	-0.119	0.110	0.154	0.475
4	-0.126	-0.027	0.021	0.029	-0.044	0.082	0.129	0.526
5(high)	0.024	0.002	-0.018	-0.058	-0.129	-0.153	0.115	0.184
$5-1$	-0.155	0.019	-0.010	-0.042	-0.007			
std err	0.113	0.077	0.075	0.081	0.115			
p-value	0.171	0.803	0.898	0.601	0.952			

Panel B: Monthly Portfolio Alphas (in percent): Mature ETFs

	Portfolios Sorted on *ShareChange*							
Portfolios Sorted on *BidAskSpread*	1(low)	2	3	4	5(high)	$5-1$	std err	p-value
1(low)	0.596	-0.148	0.050	-0.295	-0.454	-1.050	0.422	0.014
2	0.709	-0.557	-0.379	-0.290	-0.932	-1.641	0.503	0.001
3	-0.296	-0.026	0.267	-0.254	-0.955	-0.659	0.534	0.219
4	-0.371	-0.136	0.048	-0.170	-0.994	-0.623	0.448	0.166
5(high)	-0.049	-0.266	-0.341	-0.121	-0.475	-0.426	0.453	0.348
$5-1$	-0.645	-0.117	-0.391	0.174	-0.021			
std err	0.421	0.313	0.293	0.279	0.454			
p-value	0.128	0.709	0.183	0.535	0.963			

Table 7: Value-weighted portfolio alphas based on ETF arbitrage activity and bid-ask spread. ETFs are first sorted, per period (week or month), into five portfolios based on the prior period's bid-ask spread. ETFs are then sorted, within each bid-ask-spread portfolio, based on the prior period's aggregate creation / redemption activity, resulting in twenty five total portfolios. Portfolio returns are calculated based on all ETF-period observations, and these portfolio returns are regressed on the market, small-minus-big, high-minus-low and momentum factors. The last three columns and last three rows of each panel present results from testing the difference in returns between the highest and lowest portfolios.

Panel A: Weekly Portfolio Alphas (in percent): Mature ETFs

Portfolios Sorted on *BidAskSpread*	Portfolios Sorted on *ShareChange*: 1(low)	2	3	4	5(high)	$5-1$	std err	p-value
1(low)	0.074	0.043	-0.041	0.039	-0.010	-0.084	0.079	0.286
2	0.213	0.033	0.013	-0.009	-0.154	-0.366	0.116	0.002
3	-0.022	0.033	0.033	0.019	0.122	0.145	0.148	0.329
4	0.019	-0.000	0.043	0.017	0.050	0.031	0.140	0.827
5(high)	0.061	0.013	0.028	-0.026	0.053	-0.008	0.146	0.957
$5-1$	-0.013	-0.030	0.069	-0.065	0.064			
std err	0.105	0.090	0.079	0.093	0.128			
p-value	0.901	0.740	0.384	0.483	0.620			

Panel B: Monthly Portfolio Alphas (in percent): Mature ETFs

Portfolios Sorted on *BidAskSpread*	Portfolios Sorted on *ShareChange*: 1(low)	2	3	4	5(high)	$5-1$	std err	p-value
1(low)	0.292	0.071	-0.018	-0.174	-0.236	-0.528	0.305	0.085
2	0.441	-0.028	0.190	-0.214	-0.353	-0.794	0.471	0.094
3	-0.344	0.287	0.374	0.071	0.211	0.554	0.506	0.274
4	-0.392	0.179	0.160	0.037	-0.212	0.180	0.464	0.699
5(high)	0.123	-0.319	0.066	0.132	0.155	0.032	0.495	0.949
$5-1$	-0.169	-0.389	0.084	0.306	0.391			
std err	0.374	0.381	0.336	0.319	0.445			
p-value	0.653	0.308	0.804	0.338	0.380			

Table 8: Equal-weighted portfolio alphas based on ETF arbitrage activity and volume. ETFs are first sorted, per period (week or month), into five portfolios based on the prior period's volume. ETFs are then sorted, within each volume portfolio, based on the prior period's aggregate creation / redemption activity, resulting in twenty five total portfolios. Portfolio returns are calculated based on all ETF-period observations, and these portfolio returns are regressed on the market, small-minus-big, high-minus-low and momentum factors. The last three columns and last three rows of each panel present results from testing the difference in returns between the highest and lowest portfolios.

Panel A: Weekly Portfolio Alphas (in percent): Mature ETFs

	Portfolios Sorted on *ShareChange*							
Portfolios Sorted on *Volume*	1(low)	2	3	4	5(high)	$5-1$	std err	p-value
1(low)	-0.051	-0.092	-0.039	-0.083	-0.086	-0.035	0.104	0.736
2	-0.115	-0.005	0.047	-0.075	-0.128	-0.012	0.093	0.894
3	-0.076	-0.085	-0.056	-0.031	-0.250	-0.174	0.118	0.142
4	0.015	-0.025	0.054	-0.094	-0.121	-0.137	0.132	0.302
5(high)	0.059	-0.016	-0.066	-0.007	-0.081	-0.139	0.126	0.267
$5-1$	0.110	0.076	-0.027	0.076	0.006			
std err	0.109	0.083	0.082	0.076	0.121			
p-value	0.312	0.359	0.746	0.314	0.962			

Panel B: Monthly Portfolio Alphas (in percent): Mature ETFs

	Portfolios Sorted on *ShareChange*							
Portfolios Sorted on *Volume*	1(low)	2	3	4	5(high)	$5-1$	std err	p-value
1(low)	0.241	-0.518	-0.573	-0.418	-0.786	-1.027	0.385	0.008
2	-0.516	-0.454	0.205	-0.495	-0.752	-0.236	0.411	0.566
3	-0.226	-0.604	-0.091	-0.397	-1.031	-0.804	0.413	0.053
4	0.321	-0.163	-0.097	-0.229	-1.036	-1.357	0.543	0.013
5(high)	0.349	0.119	-0.170	-0.294	-0.484	-0.834	0.418	0.047
$5-1$	0.108	0.637	0.402	0.124	0.302			
std err	0.397	0.297	0.320	0.281	0.406			
p-value	0.785	0.033	0.210	0.658	0.459			

Table 9: Value-weighted portfolio alphas based on ETF arbitrage activity and volume. ETFs are first sorted, per period (week or month), into five portfolios based on the prior period's volume. ETFs are then sorted, within each volume portfolio, based on the prior period's aggregate creation / redemption activity, resulting in twenty five total portfolios. Portfolio returns are calculated based on all ETF-period observations, and these portfolio returns are regressed on the market, small-minus-big, high-minus-low and momentum factors. The last three columns and last three rows of each panel present results from testing the difference in returns between the highest and lowest portfolios.

Panel A: Weekly Portfolio Alphas (in percent): Mature ETFs

	Portfolios Sorted on *ShareChange*							
Portfolios Sorted on *Volume*	1(low)	2	3	4	5(high)	$5-1$	std err	p-value
1(low)	0.107	-0.013	0.004	0.011	-0.002	-0.110	0.132	0.406
2	0.101	-0.001	0.082	-0.027	0.021	-0.081	0.109	0.461
3	0.053	-0.029	0.042	-0.018	-0.117	-0.170	0.118	0.150
4	0.113	0.006	0.059	-0.013	-0.056	-0.170	0.098	0.085
5(high)	0.064	0.032	-0.114	0.136	0.056	-0.008	0.107	0.941
$5-1$	-0.044	0.045	-0.118	0.125	0.058			
std err	0.128	0.079	0.075	0.073	0.111			
p-value	0.733	0.566	0.115	0.085	0.601			

Panel B: Monthly Portfolio Alphas (in percent): Mature ETFs

	Portfolios Sorted on *ShareChange*							
Portfolios Sorted on *Volume*	1(low)	2	3	4	5(high)	$5-1$	std err	p-value
1(low)	1.183	-0.280	-0.151	0.096	-0.259	-1.443	0.481	0.003
2	-0.153	-0.286	0.221	-0.256	-0.006	0.147	0.416	0.725
3	-0.248	-0.072	0.171	-0.057	-0.320	-0.071	0.447	0.873
4	0.118	0.144	-0.065	0.165	-0.356	-0.474	0.417	0.257
5(high)	0.199	-0.005	-0.078	0.002	-0.541	-0.740	0.374	0.049
$5-1$	-0.984	0.275	0.073	-0.094	-0.281			
std err	0.475	0.282	0.271	0.288	0.382			
p-value	0.039	0.331	0.788	0.744	0.462			

Table 10: Distribution of share-growth-adjusted return t-stats for the Mature ETF sample using Monte Carlo simulations. Each test statistic is calculated using the simulated distribution of share-growth-adjusted returns from 100,000 Monte Carlo paths. The data set begins in January 2007 and ends in February 2016. All ETFs in the sample require 36 months of data. Panel A presents the equal weighted results, in which the percent of funds with t-stats in the respective bucket are displayed. Panel B presents the value weighted results based on market capitalization. Market capitalization weights are computed as $\sum_{T} \frac{MC_{j,t}}{\sum_{J} MC_{k,t}}/T$. The market capitalization weights are an average market capitalization share across time to control for time trends.

Active ETFs - All	
Panel A: Equal Weighted T-Stat Range	**Percent Of Sample**
$(\infty, -1.96)$	9.15%
$[-1.96, 0)$	40.38%
$(0, 1.96]$	45.07%
$(1.96, \infty)$	5.40%
Total	100.00%
Panel B: Value Weighted T-Stat Range	**Percent Of Sample**
$(\infty, -1.96)$	17.97%
$[-1.96, 0)$	44.46%
$(0, 1.96]$	32.80%
$(1.96, \infty)$	4.76%
Total	100.00%
$N = 426$	

Table 11: Distribution of share-growth-adjusted return t-stats for the Mature ETF sample using Monte Carlo simulations, funds existing prior to January 2007. Each test statistic is calculated using the simulated distribution of share-growth-adjusted returns from 100,000 Monte Carlo paths. The data set begins in January 2007 and ends in February 2016. All ETFs in the sample require 36 months of data and must be considered a mature ETF prior to January 2007. Panel A presents the equal weighted results, in which the percent of funds with t-stats in the respective bucket are displayed. Panel B presents the value weighted results based on market capitalization. Market capitalization weights are computed as $\sum_T \frac{MC_{j,t}}{\sum_J MC_{k,t}}/T$. The market capitalization weights are an average market capitalization share across time to control for time trends.

Active ETFs - Introduced Pre-2007	
Panel A: Equal Weighted T-Stat Range	**Percent Of Sample**
$(\infty, -1.96)$	7.34%
$[-1.96, 0)$	49.54%
$(0, 1.96]$	33.94%
$(1.96, \infty)$	9.17%
Total	100.00%
Panel B: Value Weighted T-Stat Range	**Percent Of Sample**
$(\infty, -1.96)$	21.13%
$[-1.96, 0)$	49.18%
$(0, 1.96]$	26.30%
$(1.96, \infty)$	3.39%
Total	100.00%
$N = 109$	

Table 12: Distribution of share-growth-adjusted return t-stats for the Mature ETF sample using Monte Carlo simulations, funds introduced after January 2007. Each test statistic is calculated using the simulated distribution of share-growth-adjusted returns from 100,000 Monte Carlo paths. The data set begins in January 2007 and ends in February 2016. All ETFs in the sample require 36 months of data and must be considered a mature ETF after January 2007. Panel A presents the equal weighted results, in which the percent of funds with t-stats in the respective bucket are displayed. Panel B presents the value weighted results based on market capitalization. Market capitalization weights are computed as $\sum_{T} \frac{MC_{j,t}}{\sum_{J} MC_{k,t}} / T$. The market capitalization weights are an average market capitalization share across time to control for time trends.

Mature ETFs - Introduced Post-2007	
Panel A: Equal Weighted T-Stat Range	**Percent Of Sample**
$(\infty, -1.96)$	9.78%
$[-1.96, 0)$	37.22%
$(0, 1.96]$	48.90%
$(1.96, \infty)$	4.10%
Total	100.00%
Panel B: Value Weighted T-Stat Range	**Percent Of Sample**
$(\infty, -1.96)$	12.26%
$[-1.96, 0)$	33.62%
$(0, 1.96]$	46.42%
$(1.96, \infty)$	7.70%
Total	100.00%
$N = 317$	

Table 13: Distribution of share-growth-adjusted return t-stats for the Mature ETF sample using Monte Carlo simulations, top 100 funds by average market capitalization share. Each test statistic is calculated using the simulated distribution of share-growth-adjusted returns from 100,000 Monte Carlo paths. The data set begins in January 2007 and ends in February 2016. All ETFs in the sample require 36 months of data and must be ranked in the top 100 funds based on market capitalization share. Panel A presents the equal weighted results, in which the percent of funds with t-stats in the respective bucket are displayed. Panel B presents the value weighted results based on market capitalization. Market capitalization weights are computed as $\sum_{T} \frac{MC_{j,t}}{\sum_{J} MC_{k,t}}/T$. The market capitalization weights are an average market capitalization share across time to control for time trends.

Mature ETFs - Top 100 By Market Cap	
Panel A: Equal Weighted T-Stat Range	**Percent Of Sample**
$(\infty, -1.96)$	11.00%
$[-1.96, 0)$	44.00%
$(0, 1.96]$	37.00%
$(1.96, \infty)$	8.00%
Total	100.00%
Panel B: Value Weighted T-Stat Range	**Percent Of Sample**
$(\infty, -1.96)$	19.87%
$[-1.96, 0)$	45.62%
$(0, 1.96]$	29.73%
$(1.96, \infty)$	4.78%
Total	100.00%
$N = 100$	